DOCUMENT RESUME

ED 094 719　　　　　　　　　　　　　　IR 000 915

AUTHOR	Cohen, Malcolm S.
TITLE	On the Feasibility of a Labor Market Information System. Volume 3. Final Report for Period July 1, 1970-June 30, 1974.
INSTITUTION	Michigan Univ., Ann Arbor. Inst. of Labor and Industrial Relations.
SPONS AGENCY	Manpower Administration (DOL), Washington, D.C. Office of Research and Development.
REPORT NO	DLMA-71-24-70-02-3
PUB DATE	Jun 74
NOTE	177p.; For related documents see IR 000 913 and 914
AVAILABLE FROM	Institute of Labor and Industrial Relations, Publications Office, 401 South Fourth Street, Ann Arbor, Michigan 48106 ($3.00)
EDRS PRICE	MF-$0.75 HC-$9.00 PLUS POSTAGE
DESCRIPTORS	*Computer Oriented Programs; *Feasibility Studies; *Information Retrieval; *Information Services; Information Theory; *Manpower Utilization; Programing Languages
IDENTIFIERS	Benchmarking; *Labor Market Information System

ABSTRACT

Volume 3 contains supporting appendixes of a three-volume study designed to evaluate the need for new labor market information. The study also sought to improve methods for making better use of currently collected statistical and administrative data. The appendixes support Volume 1's chapter 2 "A Flexible Data Retrieval System" and chapter 3 "Data Access and Quality." Volume 3 appendixes are: (1) improvements in benchmarking, (2) automated graphics for benchmarking, (3) a manpower information service, (4) a set theoretic data structure and retrieval language, (5) MICRO information retrieval system technical reference manual, (6) the LMIS (Labor Market Information Service, (7) the needs of users of a labor market information system, and (8) minority employment data sources. (WCM)

ED 094719

Report DLMA 71-24-70-02-3

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ON THE FEASIBILITY OF A LABOR MARKET INFORMATION SYSTEM

Volume III

Malcolm S. Cohen

Institute of Labor and Industrial Relations
The University of Michigan-Wayne State University
Ann Arbor, Michigan 48104

June 1974

Final Report for Period July 1, 1970-June 30, 1974

Prepared for

U.S. Department of Labor
Manpower Administration
Office of Research and Development
Washington, D.C. 20210

IR 000 915

This report was prepared for the Manpower Administration, U.S. Department of Labor, under research and development contract No. DL 71-24-70-02. Since contractors conducting research and development projects under Government sponsorship are encouraged to express their own judgement freely, this report does not necessarily represent the official policy of the Department of Labor. The contractor is solely responsible for the contents of this report.

Copyright (c) 1974 by the Institute of Labor and Industrial Relations, The University of Michigan-Wayne State University

Reproduction in whole or in part permitted for any purpose of the United States Government.

ADDITIONAL COPIES MAY BE OBTAINED FROM

Institute of Labor and Industrial Relations
Publications Office
401 S. 4th Street
Ann Arbor, Michigan 48106

This report is available in three volumes. The price is $3.00 per volume.

Order from: The Institute of Labor and Industrial Relations
Publications Office
P.O. Box B-1
University of Michigan
Ann Arbor, Michigan 48106

TABLE OF CONTENTS

Volume III

Appendix   Page

E IMPROVEMENTS IN BENCHMARKING

1.0 Introduction 1
1.1 MESC Benchmarking Procedure 1
1.2 An Automated Benchmark Worksheet 3
1.3 Statistical Methods of Error Detection 3
1.4 Proposed Models 5
1.5 Criteria for Error Detection 8
1.6 Summary of Steps Necessary to Evaluate Firm Employment Data 10
1.7 Practical Application of the Methods 11

F AUTOMATED GRAPHICS FOR BENCHMARKING

2.0 Introduction 13

F-1

2.1 Brief Description of PLOT 21
2.2 Operation of PLOT 21
2.3 Compiling the Deck of Cards 21
2.4 FORMAT Specification 25
2.5 Running PLOT and Examining Tabulations 26
2.6 Producing the CALCOMP Graphs 27
2.7 Final Procedures 28

G A MANPOWER INFORMATION SERVICE

3.0 Introduction 29
3.1 Functional Location of the Manpower Information Service (MIS) 29
3.2 Relation of MIS to State Employment Agencies 30
3.3 Internal Structure of the Manpower Information Service 30
3.4 Personnel of the MIS 34

TABLE OF CONTENTS (Continued)

Volume III

Appendix Page

G 3.5 Overall Budget Parameters 37

3.6 Conclusion 38

H A SET THEORETIC DATA STRUCTURE AND RETRIEVAL LANGUAGE

4.0 Introduction 39

4.1 General Organization and Function of the Program 40

4.2 Structure and Content of the Data Bases 44

4.3 Use of the Program and Data-Bases 44

4.4 Program Control and Use of the Data-Structure 46

4.5 Directory 47

4.6 The Dictionary 50

4.7 The Data-Set 53

4.8 Limitations and Plans for Improvement 54

I MICRO INFORMATION RETRIEVAL SYSTEM (VERSION 3.9)
TECHNICAL REFERENCE MANUAL

5.0 System Basics 57

5.0.1 Introduction 57

5.0.2 Concepts and Facilities 57

5.0.3 General Information 60

5.0.4 Security Features of MICRO 62

5.1 Command Descriptions 64

5.2 Macro Subsystem 95

5.2.1 Introduction 95

5.2.2 Macro Libraries 95

5.2.3 Macro-Definitions 95

5.2.4 Macro-Command 96

5.2.5 Variable Symbols 96

5.2.6 Writing Macro-Descriptions/Commands 97

I-1 MIDAS 104

TABLE OF CONTENTS (Continued)

Volume III

Appendix | Page

I-2 MACRO EXAMPLE 105

J THE LMIS DATA BASE VERSION 2
- 6.0 Introduction 107
- 6.1 Census of Population 107
- 6.2 Current Population Survey 108
- 6.3 EEO-1 108
- 6.4 ESARS 109
- 6.5 ES202 109
- 6.6 ES203 110
- 6.7 Job Banks 110
- 6.8 Social Security 111
- 6.9 Urban Employment Survey 111

J-1 DATA SETS IN THE LMIS DATA BASE 112

K MEETING THE NEEDS OF USERS OF A LABOR MARKET INFORMATION SYSTEM
- 7.0 Introduction 115
- 7.1 Needs of Users 115
- 7.2 The Users And Their Needs 116
- 7.3 Management Information 125

L AN INDEX TO MAJOR PUBLISHED DATA ELEMENTS FOR USERS OF LABOR MARKET INFORMATION
- 8.0 Introduction 127
- 8.1 Sources 127

M MINORITY EMPLOYMENT DATA SOURCES FOR SMSA'S
- 9.0 Introduction 145
- 9.1 Milwaukee 146

TABLE OF CONTENTS (Continued)

Volume III

Appendix Page

M 9.2 Denver and Detroit 148

9.3 Previous Tabulations 150

9.4 Conclusions and Recommendations 150

REFERENCES 169

LIST OF FIGURES

Volume III

FIGURE PAGE

1 Time Series Plots - UI Data vs. Predicted Employment 7
2 Computerized Plotting To Do Benchmarking 15
3 Multiple Plots From the CALCOMP 16
4 Plot Program Numeric Output 17
5 Plot Program 19
6 Manpower Information Service 32
7 Contrasting Data Structures 42
8 MICRO Interface To Set Theoretic Operations 43
9 Directory Entry Formation Disk 48
10 Directory Format in Core 48
11 MICRO Combines Dictionary and Directory 49
12 Data Set Control Block (DSCB) 51
13 The R Block 52
14 The S Block 52

LIST OF TABLES

Volume III

TABLE		PAGE
1	How Jobs Are Found	118
2	Source of Job Guidance	120
3	Demographic Characteristics	137
4	Employment	139
5	Wages	140
6	Geographic Breakdown	141
7	Labor Supply	143
8	Milwaukee SMSA, 1970 - Comparison Between 20% Sample of the Census and 2% Sample	153
9	Milwaukee SMSA, 1970 - Comparison Between EEOC and Census 2% Sample	154
10	Milwaukee SMSA, 1970 - Analysis of Employment by Size of Firm by Industry	155
11	Milwaukee SMSA, 1970 - Employment Industry by Occupation - Comparison Between EEOC and Census 2% Sample	156
12	Milwaukee SMSA, 1970 - Comparison Between Census 2% and EEOC Industry by Occupation	157
13	Denver SMSA, 1970 - Comparison Between EEOC and Census 2% Sample	159
14	Denver SMSA, 1970 - Employment Industry by Occupation - Comparison Between EEOC and Census 2% Sample	160
15	Denver, SMSA, 1970 - Comparison Between Census 2% Sample and EEOC Industry by Occupation	161
16	Detroit SMSA, 1970 - Comparison Between EEOC and Census 2% Sample	163
17	Detroit SMSA, 1970 - Employment Industry by Occupation - Comparison Between EEOC and Census 2% Sample	164
18	Detroit SMSA, 1970 - Comparison Between Census 2% and EEOC Industry by Occupation	165
19	Minority Group Employment by Occupation and Sex For Selected Industries and Standard Metropolitan Statistical Areas, 1967	167

APPENDIX E

IMPROVEMENTS IN BENCHMARKING

Stephen T. Marston*

1.0 Introduction

The goal of the benchmarking project is to design statistical procedures and specifications for improving the quality and timeliness of encountered data bases, as well as to assess the costs of doing so. An encountered data base results from administrative data collected as part of a legal requirement, program requirement or an ongoing statistical program. The specific experiment uses ES202 and BLS 790 data for Detroit, Michigan. ES202 provides employment data and wage data by quarter for every firm covered by the Unemployment Insurance System. The BLS 790 program provides monthly hours, wages and employment data on a sample of firms.

This report reviews current procedures in benchmarking and suggests some improvements and methods of automating these procedures. Starting with section 1.4, it presents a new method for improving the quality of data.

1.1 MESC Benchmarking Procedure

Currently the Michigan Employment Security Commission (MESC) prepares one benchmark estimate each year for submission to the Manpower Administration in April. The benchmark is MESC's best estimate of employment in firms and industries during March of the previous year. It is calculated by correcting ES202 (UI) employment data of March for errors, omissions and misclassifications. The new benchmark is then used to adjust BLS 790 (BLS) industry employment data for later months by multiplying those figures by the ratio between the benchmark employment for the industry and the previous estimate of employment for the industry. This makes the assumption that the magnitude of errors in later months of the year are proportional to the magnitude of errors in March. The employment data for the twelve months between the last benchmark and the current benchmark are adjusted by a factor which is interpolated linearly between the two benchmarks.

---

* The Labor Market Information Systems Project is sponsored by the Manpower Administration, Office of Research and Development, United States Department of Labor, under contract no. 71-24-70-02. The views represented in this paper are the sole responsibility of the author and do not necessarily reflect the views of the Department of Labor.

The principal method used by MESC to search out errors in the March employment data is a firm-by-firm comparison of UI employment with BLS employment. If it is found that the employment for the firm is substantially different in BLS that in UI data or that the firm is classified in a different industry in BLS than in UI data, then further research is required to establish the correct employment or classification. This procedure is repeated for all firms of more than fifty workers for which both UI and BLS data exists. Since UI data includes all firms covered by unemployment insurance and BLS data includes only a sample of those firms, there will be firms for which UI data is available, but BLS data is not available.

Ajeet Kang and the author have reviewed the procedures now followed at MESC for benchmarking. The methods could be improved and result in considerable cost savings of at least one man-year for example, in Michigan.

The comparison of BLS with UI data by firm is accomplished in an entirely manual way. Tabulations of BLS and UI firm data are categorized in folders and then manually paired, UI employment for the firm with BLS employment for the firm. This process, which must be finished prior to the actual data checking, may hold up the benchmarking and occupy valuable employees for a month.

Furthermore, the tabulations themselves are always a source of difficulty. They are often late in arrival; the 1971 tabulations will be several months behind schedule. More important is the fact that the present tabulations are not specially designed for benchmarking and are, in fact, largely unsuitable for benchmarking. There are several reasons for this:

The BLS tabulations use a different labor market area code than do the UI tabulations. Thus the two data sets can not be compared without a translation table. This simple inconsistency is responsible for a substantial labor waste and error.

2) The BLS tabulations do not include the employer number as do UI tabulations. Identifying a firm from the UI tabulations on the BLS tabulations might require reference to the firm name (which is not unique), the firm address and the report number. This is slow and can be inexact.

3) The BLS tabulations contain only one month of employment data. Frequently more months are needed to make an informed analysis. At least three months should be on the tabulations; twelve months would be better still.

4) BLS tabulations should include the name of the firm.

1.2 An Automated Benchmark Worksheet

Therefore the present BLS and UI tabulations must be reprogrammed. The changes must include the above suggestions and the output must have the approval of the benchmark preparer at MESC. The new tabulations must be created regardless of whether further suggestions are followed.

A benchmark worksheet must be programmed. The program which produces the worksheet should select firms included in both the UI and BLS data in which the employment figures differ substantially. The program should print out data on such firms when the employment is greater than some specified level. This worksheet should then be used, in connection with the above tabulations, to prepare the benchmark.

When a worksheet is needed, the program should be run by the benchmark preparer himself. This will be more likely to produce the worksheets when they are needed by the benchmark preparer. This individual should also be free to make any changes he desires in the program itself.

1.3 Statistical Methods of Error Detection

The UI/BLS comparison points out errors occurring in only one tabulations or the other, but not both. This is sufficient for catching a wide range of clerical errors and industry misclassifications. However, some errors will occur in both UI and BLS data and will cause them to agree and the errors not to be caught by the above procedure. In order to bring more information into this data testing, the Labor Market Information System Project (LMIS) proposes to predict firm employment using statistical techniques. The predicted employment can then be compared with the UI employment. If they differ substantially, as in the case of the UI/BLS comparison, further investigation of the data is indicated to the MESC analyst.

The value of the statistical technique lies in summarizing information about the seasonal nature of the firm, past employment levels and employment of the whole industry to see whether the UI employment is extraordinary by comparison with past behavior of the firm. This requires building an economic model which incorporates these factors to explain the employment of the firm.

The design of this model is tightly confined by the data and the benchmarking process itself:

1) A separate model must be estimated for each firm, but all models must contain the same variables. In a precise econometric analysis different variables would be found to influence the employment of different firms and an individual decision would have to be made on the model for each firm. That is clearly impossible in this case where thousands of firms are involved. This limits the models to generally applicable equations rather than precise and specific ones.

2) Time-series data on firm employment comes at great cost in programmer time and effort to the person using the procedure developed. Only three consecutive months of data are included on each UI tape. In order to acquire data going back more than three months other tapes must be mounted and the data sorted so that employment for one firm is connected with employment for the same firm in different months.

3) Even if long series of historical data are collected they may be meaningless. The firm may have changed location, changed its industry, or changed it business behavior. Data more than a year old is more likely to represent a substantially different firm than data only six months old.

1.4 Proposed Models

The author has been experimenting with models of the following two forms:

$$E_{it} = b_{0i} + b_{1i}E_{i,t-1} + b_{2i}L_{jt} \qquad (1)$$

and

$$E_{it} = b_{0i} + b_{1i}E_{i,t-1} + b_{2i}L_{jt} + b_{3i}t \qquad (2)$$

where E_{it} = employment of firm i in month t

L_{jt} = employment in industry j of which firm i is a member excluding firm i itself

$$= \sum_{k\in j} E_{kt} - E_{it} = \sum_{\substack{k\in j \\ k\neq i}} E_{kt}$$

The first equation models firm employment as a function of last month's employment and the employment in the whole industry. The employment of firm i itself must be excluded from the industry employment so as not to tautologically make the firm's employment a function of itself. The second model allows for a trend factor by adding time as an independent variable.

The industry employment variable represents the behavior of other firms in the industry. Factors such as seasonal influences and short-term changes in product demand should affect the employment of all firms in an industry. Thus employment of a particular firm should generally conform to the changes occurring in the employment of the rest of the industry. Industry employment serves as a proxy for short-term changes in product demand.

The author has pieced together a twelve month time series of firm employment and industry employment from Detroit UI data for 1970. From these data he has estimated equations (1) and (2) for each of the firms. Equation (2) performs marginally better than equation (1). Example statistics from the regressions based on equation (2) are as follows:

Firm	Industry (SIC)	R^2	Partial Correlation Coefficients b_1	b_2	b_3
1	64	0.46	.07	-.55	.05
2	42	0.78	.43	.85	-.29
3	34	0.89	.92	.84	.87
4	37	0.40	-.01	.50	-.06

The four regressions differ considerably in their predictive power. Regressions for firms 2 and 3 are about twice as good predictors as are regressions for firms 1 and 4. The coefficient b_2 for firm 1 is slightly disturbing. The negative sign indicates that the firm operates on a cycle counter to that of the rest of the industry. Perhaps that is because this firm picks up some of the demand the other firms lose and vice versa.

The regressions explain enough variance that they can be used for error checking. The method is simply to isolate outliers. To do so the author has graphed observed firm employment against predicted firm employment for each firm. Figure 1 contains four examples of such graphs.

Firms 1, 2 and 3 are similar: observed and predicted employment run roughly in the same time pattern. There appears to be no reason to question any particular month in the graph. However firm 4 displays an irregularity in month 9. Firm 4 employment in September is about 700 employees more than is predicted by the model. This indicates an outlier and so the employment of firm 4 should be investigated to find out if there is a reason why employment should be so large in September of 1970.

If this process were repeated on a regular basis it would substantially improve the quality of the UI employment data. However it may not be an acceptable procedure for error checking among the tens of thousands of firms in Michigan.

1) It is still too "manual" and slow. Too many graphs must be searched carefully by human eye even if the graphs themselves are produced by computer.

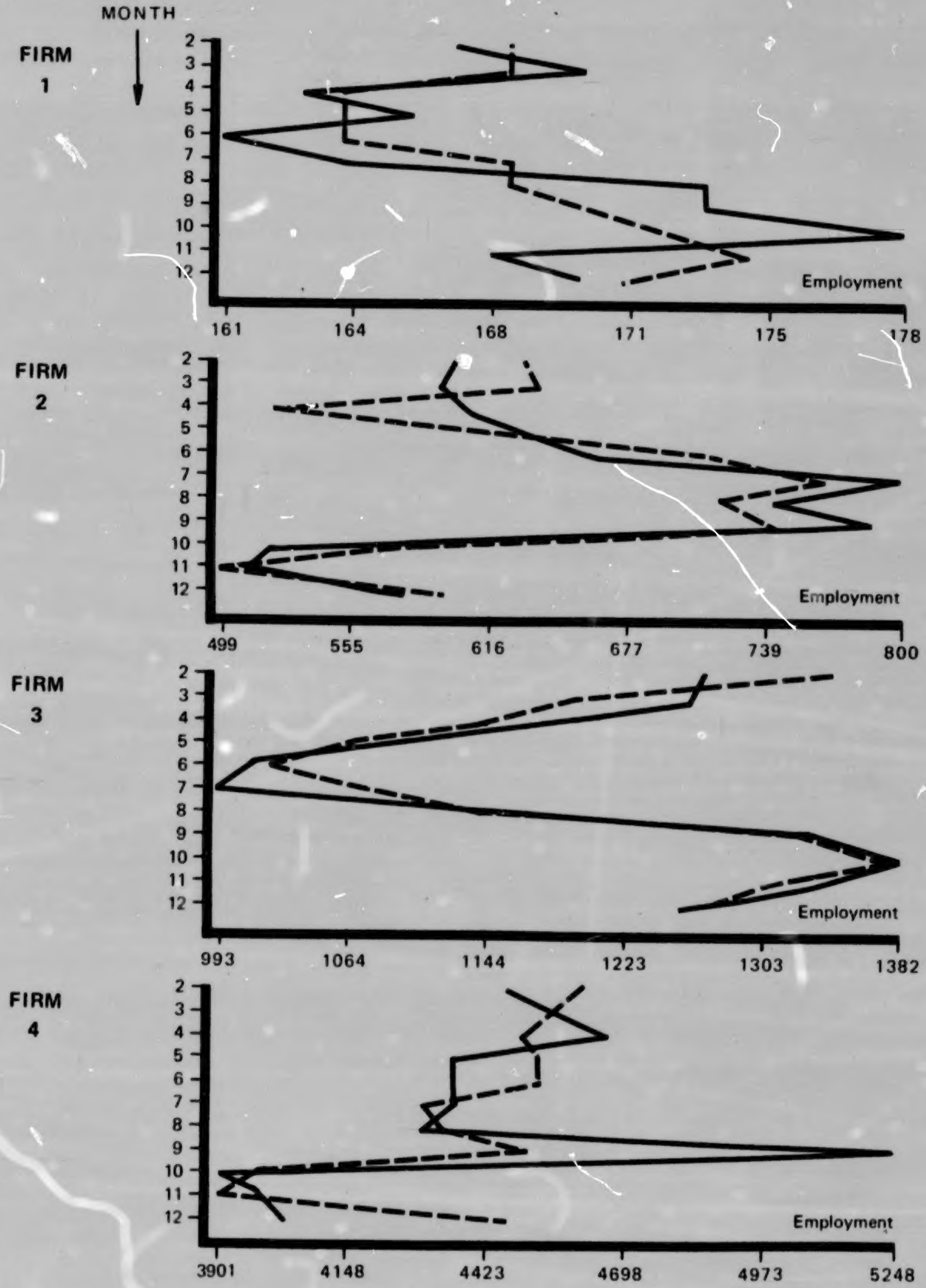


Figure 1 Time Series Plots UI Data (–) VS. Predicted Employment (---)

2) There is no objective criterion upon which to choose outliers. Presumably different people might disagree as to whether a particular point on a graph is an outlier. Furthermore the best possible criterion must be chosen in order to catch the largest number of errors, while avoiding the extra work of labeling data points errors, only to find out later that they are correct.

1.5 Criteria for Error Detection

Consider first the simple criterion that a particular data point is an outlier if the residual between the predicted and observed employment is greater than some constant multiplied by the standard error of the regression (SE). The SE is the standard deviation of the residuals from the regression and can be calculated by

$$SE^2 = \frac{1}{n-k} \sum_{t=1}^{n} (\hat{E}_t - E_t)^2 \qquad (3)$$

where n = number of months

k = number of estimated parameters in regression

E_t = employment in month t

$\hat{E}_t$ = predicted employment in month t

This criterion would pick out the largest residuals as outliers. Furthermore since it is defined with respect to the residuals themselves it would only pick out a particular residual as an outlier if it was significantly greater than the other residuals. This has the advantage and disadvantage that it automatically adjusts the criterion for the case of a bad fitting regression. A bad fitting regression will have a large SE, and a residual may be relatively large without failing to pass the criterion. Likewise a regression which fits well will have a small SE and the criterion will be relatively more binding. Thus not just large residuals would be caught, but residuals large relative to other residuals from the same regression. The absolute size of the residuals would be irrelevant, and the closeness of the fit of the regression would be irrelevant.

It would be proper to use a criterion which does not depend upon the closeness of the fit of the regression if there were no reason *a priori* to believe that the fit should be close. In that case only the *a posteriori* information that the data of a firm fit the model well would be evidence that the model is good for that firm. If the fit is not good it could be because the model simply is not good for the firm and does not indicate a likely error in the data.

However if there were reason to have faith in the basic correctness of the model, then a bad fit (a low R^2) would call the data into question. This the criterion we are considering does not do. It is completely independent of the coefficient of determination (R^2) of the regression. This seems unnecessarily extreme; the R^2 should be considered.

To what extent should the R^2 be considered? This question can be answered by discriminant analysis. Discriminant analysis can be used to calculate a discriminant function f which will divide the data into two groups: one likely to be in error (G_1) and one not likely to be in error (G_2).

Let

R_i^2 = coefficient of determination of the regression for the ith firm

SE_i = standard error of the regression for the ith firm

e_{it} = the residual from the regression for the ith firm at month t.

Then e_{it}/SE_i is the relative error of observation t in the regression for firm i.

The discriminant function, f_{it}, is a linear combination of R_i^2 and e_{it}/SE_i which partitions the data into an error group, G_1, and a correct group, G_2.

$$f_{it} = a_0 + a_1 R_i^2 + a_2(e_{it}/SE_i) \qquad (4)$$

where $a_2 > 0$ and $a_1 < 0$

The classification rule is

$$E_{it} \text{ is in } \begin{pmatrix} G_1 \\ G_2 \end{pmatrix} \text{ if } \begin{pmatrix} f_{it} > 0 \\ f_{it} < 0 \end{pmatrix}$$

In this application discriminant analysis consists of estimating the three weighting parameters a_0, a_1, a_2. After the parameters have been estimated they can be used to evaluate data. The parameters need only be estimated once and then they can be used repeatedly for the benchmark.

1.6 Summary of Steps Necessary to Evaluate Firm Employment Data

1) Concatenate several quarters of employment data, each originally on a separate ES202 tape, to form a twelve- or fifteen- month time series of employment for each firm. This requires that a computer program find a particular firm on each ES202 tape and transfer three months of employment data from each such tape to a single new tape. During this process the computer program can check whether the firm has the same SIC code, labor market area, or in fact is present, in all quarters. If not, the firm data should be investigated by the preparer of the benchmark. This effort constitutes a first sifting of the data. Linking the data and checking the firms to see if there are missing firms or unequal SIC codes between quarters, can be done by a computer program. Investigating the resulting list of firms, which are suspect must be done by the benchmark preparer.

2) Calculate industry employment totals. This is easily done by computer.

3) Regress each firm's employment on the industry's employment minus the firm's employment, previous month's employment and time. Calculate the standard error of the regression, all of the residuals from the regression and the R^2. All done by computer.

4) Use the coefficients from discriminant analysis (a_0, a_1, and a_2) to form the linear combinations (f_{it}) of equation (4). If f_{it} is greater than zero, E_{it} is suspected of being in error. It must be checked by the benchmark preparer.

1.7 Practical Application of the Methods

In practice the entire process could be done with two computer programs, one to do the data linking in step 1 and one to do the other calculations. The benchmark preparer would only be concerned with investigating firms which the computer programs suggested have erroneous employment data from ES202. He would not need to know anything about discriminant analysis or even regression. All of these statistical manipulations would be done for him automatically by the computer programs.

For example the programs might conclude that firm number 9,051,300 is likely to be in error in March and April of 1973. The programs would print out useful information about that firm (like its name, address, past employment, SIC code, etc.) and possibly a graph of its employment as compared with its predicted employment. All of this information should make it possible for the benchmark preparer to find out what the difficulty is in March and April, even if he has to contact the firm itself in order to resolve the problem.

The entire method can only be considered experimental at the present time. It requires substantial developing and testing yet. Testing under a simulated benchmark situation is underway with a view toward analyzing the kinds of errors picked out by the mehtod. If it is successful it holds the possibility of substantially more accurate and automated employment data error checking.

APPENDIX F

AUTOMATED GRAPHICS FOR BENCHMARKING

Stephen T. Marston*

2.0 Introduction

An important method used by the state agencies for analyzing monthly data series is graphing on the same axes those series which are alternative estimates of the same quantities. Then the series can be easily compared by eye and major discrepancies between the series isolated. These discrepancies usually indicate the existence of errors in the data which require some investigation. The most common use of this method is in the industrial series on employment. ES202 provides a census of employment in all firms in the unemployment insurance (UI) program. BLS 790 provides an estimate of employment based on a sample of firms in the UI program. A graphical procedure for analyzing these series consists of plotting on the same axes 24 months of ES202 data, 24 months of BLS 790 data with the current benchmark, and 24 months of BLS 790 data with the previous benchmark.

Volume II of the Operating Guide for the Current Employment Statistics Program, Bureau of Labor Statistics, (10.3 - 4) states that the maintenance of such charts,

> "... is strongly recommended....As resources permit, a chart series should be maintained for each estimating cell, each published industry group, the industry divisions, and total nonagricultural employment....Regular review of the charts will often reveal developing weaknesses in the estimates before they reach serious proportions...."

Unfortunately, substantial resources are required to produce these graphs. Previously at the Michigan Employment Security Commission (MESC) a statistical clerk was paid to work full-time drawing only graphs of employment. The graphs put out by this worker were generally of low quality and the job was apparently a very disagreeable one, consisting of the most stultifying form of repetition.

---

* The Labor Market Information Systems Project is sponsored by the Manpower Administration, Office of Research and Development, United States Department of Labor, under contract no. 71-24-70-02. The views represented in this paper are the sole responsibility of the author and do not necessarily reflect the views of the Department of Labor.

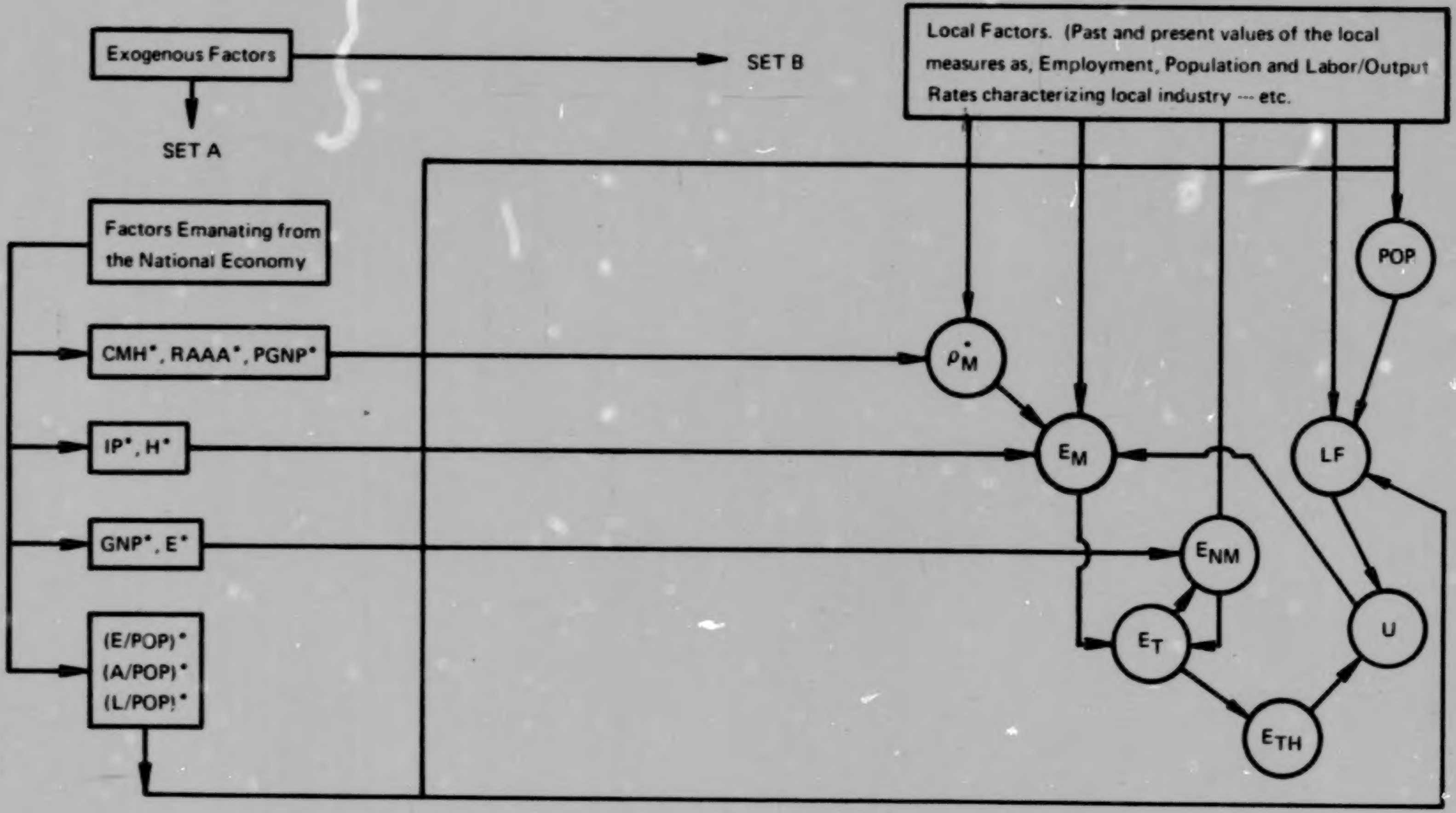
Acting upon the suggestion from MESC the Labor Market Project (LMIS) has written a computer program which will direct a CALCOMP plotter to produce all of the graphs described above and will also produce a tabulation of differences among the series plotted. The advantages of using this new program instead of the manual method are three: quality, time and cost.

The quality of the automatic graphs can be seen in Figure 2, which shows one such graph from the CALCOMP plotter. The lines are drawn crisply in black ink with the three series differentiated by three different kinds of lines: the solid line is ES202 data, the long-dashed line is BLS 790 data of old benchmark, the short-dashed line is BLS 790 data of the new benchmark. The employment scale on the vertical axis is automatically adjusted so that the data values fall conveniently within the confines of the graph. The graph is labeled according to type of data (EMPLOYMENT), SIC code (SIC 26) and month. Figure 3 shows how a large number of plots are arranged on the 31 inch wide CALCOMP paper. Figure 4 shows the tabulations which the program puts out; the first row is actual differences in employment, the second row is percentage differences. These are printed out for each industry.

The time expended in producing these graphs is allocated to two factors: keypunching time and control card arrangement. Keypunching is needed to put the data on computer cards. However, if the data has already been keypunched for some other purpose this time can be spared because the program can be adapted to most data sets. Even if the keypunching must still be done it will produce a deck of cards which may be useful for other purposes. Arranging the control cards should only take a statistical clerk an hour.

The cost of running the program is very small. Twenty-three graphs were produced at a cost of $2.40 for computer time and $2.40 for CALCOMP time, or about twenty cents per graph. This should be compared to the cost of a frustrated statistical clerk working a week or more full-time

Figure 2 Principal Forces Determining Local Employment And Unemployment In The LMIS Model.[a]


Exogenous Factors
SET A
SET B
Local Factors. (Past and present values of the local measures as, Employment, Population and Labor/Output Rates characterizing local industry --- etc.
Factors Emanating from the National Economy
CMH*, RAAA*, PGNP*
IP*, H*
GNP*, E*
(E/POP)*
(A/POP)*
(L/POP)*
POP
$\dot{\rho}_M$
E_M
LF
E_{NM}
E_T
U
E_{TH}

[a] Items with asterisks are national variables.

Figure 3 Multiple Plots From The CALCOMP

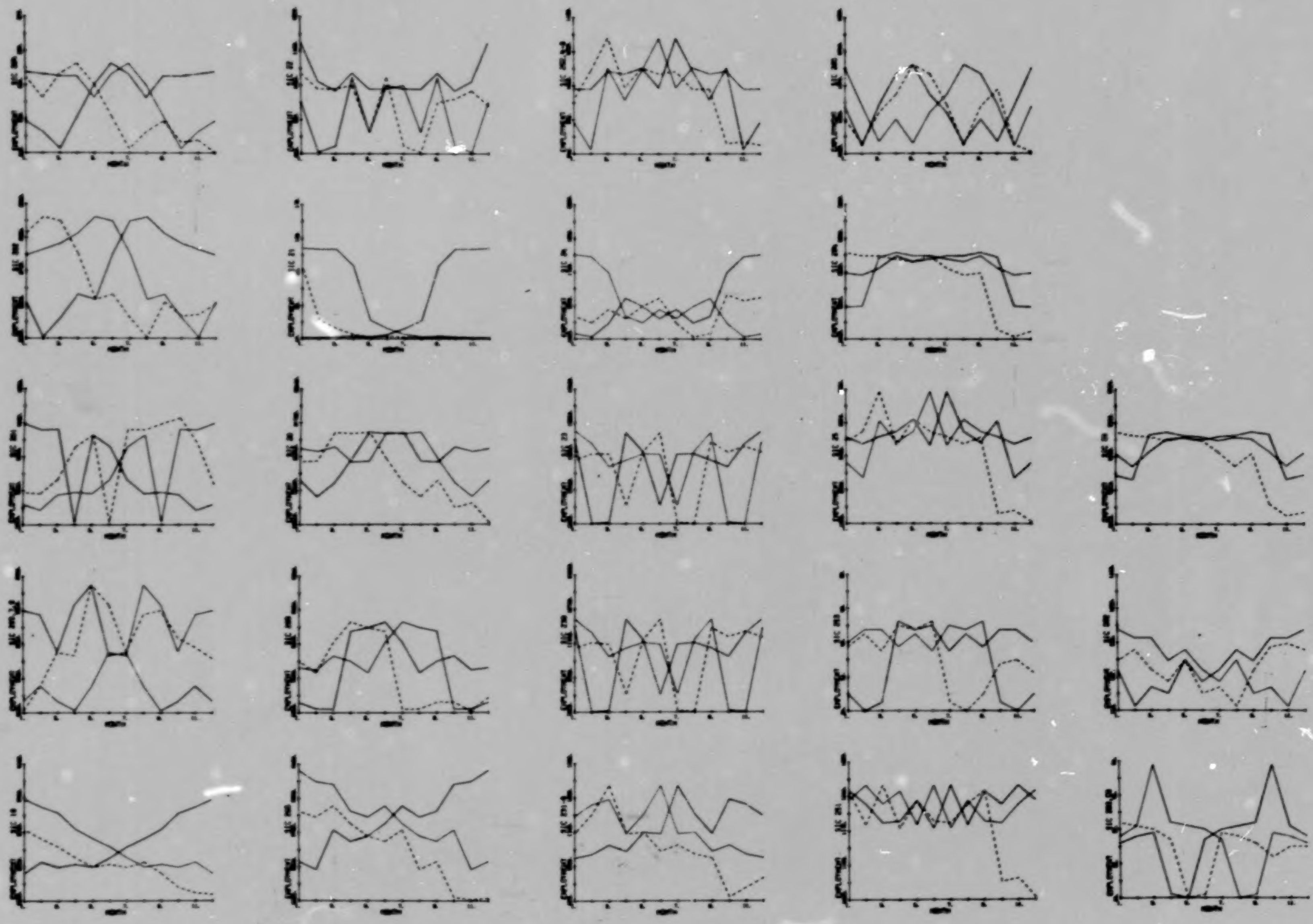


Figure 4 Plot Program Numeric Output

```
The number of plotting points specified on the control card is 12

                                   Differences
Sic Code          1           2           3           4           5
------------------------------------------------------------------------
19              -869.       -644.       -633.       -410.       -331.
             -0.2357     -0.1790     -0.1796     -0.1229     -0.1026

203,7,9          265.        209.        145.        309.        296.
              0.0859      0.0668      0.0470      0.1011      0.0946

201              190.        189.        154.        -70.        138.
              0.0480      0.0479      0.0387      0.0176      0.0346

202             -111.       -209.       -177.       -145.       -196.
             -0.0400     -0.0749      0.0631      0.0512      0.0683

204              -23.        -27.        -34.        -19.          6.
             -0.0958     -0.1130     -0.1429     -0.0798      0.0263

205             -531.       -513.       -265.       -176.       -111.
             -0.0867     -0.0847     -0.0439     -0.0299     -0.0190

208             -161.       -181.       -249.        140.        210.
             -0.0309     -0.0348     -0.0473      0.0276      0.0405

20              -371.       -532.       -426.         39.        343.
             -0.0173     -0.0249     -0.0199      0.0018      0.0161

21              -107.       -106.       -105.        -84.        -20.
             -0.9068     -0.9060     -0.8974     -0.3660     -0.6250

22              -176.       -207.       -162.        -13.       -126.
             -0.1446     -0.1880     -0.1508     -0.0115     -0.1168

231-8            -97.       -116.       -106.        -42.        -42.
             -0.0824     -0.0963     -0.0872     -0.0369     -0.0356

239             -408.      -3677.      -2520.       1341.        141.
             -0.0334     -0.3171     -0.2404      0.1232      0.0127

23              -505.      -3793.      -2626.       1299.         99.
             -0.0377     -0.2964     -0.2245      0.1081      0.0080

24              -377.       -382.       -243.         87.         79.
             -0.2081     -0.2120     -0.1407      0.0573      0.0531

252,4-9           77.        142.         -8.         61.          7.
              0.0842      0.1671     -0.0077      0.0631      0.0068

251               11.        -22.        -22.        -35.         28.
              0.0091     -0.0180     -0.0183     -0.0288      0.0238

253              -12.        -19.        -17.          6.          1.
             -0.3000     -0.4419     -0.3953      0.1538      0.0238

25                76.        101.        -47.         32.         36.
              0.0352      0.0477     -0.0206      0.0144      0.0160
```

to draw such graphs.

The most serious lmitation to this program is that it must presently be run at Wayne State University (WSU) or the University of Michigan (U-M) under the control of the Michigan Terminal System (MTS). This is convenient for MESC since it is located only one block from the WSU computing center. But it makes the program as it is presently written unavailable to other states. However, this experiment demonstrates both the cost and benefit of the method and indicates the advantage of similar projects in other states.

In this regard two approaches are possible: 1) renting time on computers and CALCOMP plotters of other private or public organizations, or 2) establishing such facilities "in house". The present project benefited greatly by using the U-M and WSU facilities. The ability of these organizations to spread the overhead cost of plotting is evident in the small cost of the plots. It seems unlikely that an owned facility would be able to match these cost estimates.

Figures 5 presents the FORTRAN IV source program which produces the tabulations and controls the plotter. This may be adapted to the use of other state organizations.

Appendix F-1 is the manual which tells how to use the program as it is presently being done at MESC.

Figure 5 Plot Program

```
      REAL MONTH(2)/'MONT','H'/,TIME(24),DIFF(24),PDIFF(24),
     *  TITLE(8)/4*'    ','SIC '/,CON(2),CNTRL(2)/'CONT','ROL '/,
     *  DATA/'DATA'/,ES202(24),BLSOLD(24),BLSNEW(24),F1(20),F2(20),
     *  F3(20),YINC/5.5/,XMX/100./
      NEND=1
      NPLOT=0
      READ(4,21)CON,N,(TITLE(J),J=1,3)
   21 FORMAT(A4,A3,I2,3A4)
      IF(ICLC(7,CON,0,CNTRL,0))25,26,25
   25 WRITE(6,28)
   28 FORMAT('0CONTROL CARD MISSING OR MISPLACED.')
      GO TO 80
   26 WRITE(6,42) N
   42 FORMAT('0THE NUMBER OF PLOTTING POINTS SPECIFIED ON THE',
     *  ' CONTROL CARD IS',I3)
      IF (N .GE. 1 .AND. N .LE. 24) GO TO 27
      WRITE(6,41)
   41 FORMAT(' IT MUST BE BETWEEN 1 AND 24.')
      GO TO 80
   27 DO 91 I=1,N
      TIME(I)=I
   91 CONTINUE
      READ(4,43)F1
   43 FORMAT(20A4)
      READ(4,43)F2
      READ(4,43)F3
      READ(4,21)D
      IF(D .EQ. DATA) GO TO 48
      WRITE(6,49)
   49 FORMAT('0DATA CARD MISSING OR MISPLACED.')
      GO TO 80
   48 CALL PGNHDR
      AXLTH=0.5*(N-1)
      XINC=AXLTH+2.5
      CALL PLTXMX(XMX)
   50 XMOVE=2.0
   30 YMOVE=1.5
      DO 40 I=1,5
      READ(5,F1,END=80)  (TITLE(J),J=6,8),(ES202(J),J=1,N)
      READ(5,F2,END=81)  (BLSOLD(J),J=1,N)
      READ(5,F3,END=81)  (BLSNEW(J),J=1,N)
      N1=1
   87 DO 82 J=N1,N
      DIFF(J)=BLSNEW(J)-ES202(J)
      PDIFF(J)=DIFF(J)/ES202(J)
      IF (J .EQ. 12) GO TO 88
```

Figure 5 Plot Program (continued)

```
82 CONTINUE
   J=N
88 WRITE(6,98)  (TITLE(K),K=6,8),(DIFF(K),K=N1,J)
98 FORMAT('0',2A4,A2,12F10.0)
   WRITE(6,99)  (PDIFF(K),K=N1,J)
99 FORMAT(11X,12F10.4)
   IF (J .EQ. N) GO TO 56
   N1=J+1
   GO TO 87
56 CALL PSCALE(4.0,0.5,EMIN,FACTOR,ES202,N,1,BLSOLD,N,1,BLSNEW,N,1)
   CALL PLTOFS(1.0,2.0,EMIN,FACTOR,XMOVE,YMOVE)
   CALL PAXFMT('F2.0')
   CALL PAXIS(XMOVE,-YMOVE,MONTH,-5,AXLTH,0.0,1.0,2.0,0.5)
   CALL PAXFMT('F6.0')
   CALL PAXIS(XMOVE,-YMOVE,TITLE,32,4.0,90.0,EMIN,FACTOR,0.5)
   CALL PLINE(TIME,ES202,N,1,0,0,1)
   CALL PDSHLN(TIME,BLSOLD,N,1,0,1)
   CALL PDSHLN(TIME,BLSNEW,N,1,0.04,1)
   NPLOT=NPLOT+1
   YMOVE=YMOVE+YINC
   NEND=0
40 CONTINUE
   XMOVE=XMOVE+XINC
   IF(XMOVE+AXLTH .LT. XMX) GO TO 30
   CALL PLTEND
   NEND=1
   GO TO 50
81 WRITE(6,46)
46 FORMAT('ODATA INCOMPLETE.')
80 IF (NEND .EQ. 0) CALL PLTEND
   WRITE(6,45)NPLOT
45 FORMAT(I15,' PLOTS GENERATED.')
   CALL SYSTEM
   END
```

APPENDIX F-1

2.1 Brief Description of PLOT

PLOT is a program for automatically graphing monthly data series and printing tabulations of differences and percentage differences between the series. It graphs three data series on each coordinate system to facilitate (for example) placing ES202 employment data, BLS 790 employment data benchmarked in 1970, and BLS 790 employment data benchmarked in 1971 on the same coordiantes. This makes it easy to compare the three series by eye. PLOT will draw any number of such graphs and will label each graph according to the need of the person running the program. Usually the PLOT user will want to draw a graph for each industry and label the graphs with their SIC codes. The number of months of data graphed can be chosen by the user up to a maximum of 24 months.

PLOT may be run at either the Wayne State Computing Center or at the University of Michigan Computing Center and operates under control of the Michigan Terminal System (MTS).

2.2 Operation of PLOT

The description of the operation of PLOT will be grouped under five headings:

1) Compiling the Deck of Cards
2) FORMAT Specification
3) Running PLOT and Examining Tabulations
4) Producing the CALCOMP Graphs
5) Final Procedures

The presentation assumes the user of PLOT knows little about MTS, however, more knowledge of MTS will clarify the reason underlying the procedures. The methods presented are correct for the version of MTS installed as of 1/73. MTS is in continuous change and may outdate anything written here.

2.3 Compiling the Deck of Cards

The initial deck of cards may be prepared anywhere there is a keypunch machine. All of the following lines should be punched starting in column 1 of an 80-column card unless otherwise indicated.

Card 1

```
$SIGNON CCID T=1M
```

This card signs the user on MTS. CCID is the user's computing center identification number. T=1M raises the maximum computer time for the job from 1/2 minute to one minute. The amount of time necessary will depend upon the number of graphs being drawn and how many months will be plotted on each graph. A recent run of PLOT creating 23 graphs of 12 months each took 17 seconds. For that run one minute would be more than adequate to complete the job. But a job drawing more graphs or more months will require proportionally more time and may take longer than one minute. For example a job drawing 230 graphs may take 170 seconds and a time specification of at least T=3M will be necessary. If the time limit is exceeded MTS will print a message saying ****Global Time Limit Exceeded**** and the user will have to resubmit his deck with a larger time specification.

Card 2

```
PASSWORD
```

This card contains only the user's computer password.

Card 3

```
$CREATE FILE SIZE=100P
```

This command creates a file for later use. The word FILE is the name of the file and may be any combination of letters and numbers the user wishes.

Card 4

```
$RUN HBG5:PLOTOBJ+*PLOTSYS 9=FILE
```

FILE is the name of the file created by card 3.

Card 5

Columns 1-7: CONTROL

Columns 8-9:	The number of months of data you want to plot in each graph "right justified". This means if the number of months is only a one digit number it should be placed in column 9. This number must be between 2 and 24.
Columns 10-21:	The label to be written on the vertical axis of each one of the graphs. This label should tell what kind of data are being plotted.

For example EMPLOYMENT might be punched in columns 10-19 and columns 20-21 left blank. Or, EARNINGS might be punched in columns 10-17.

Card 6

This card describes how the SIC codes for each industry and the first series of data to be plotted are punched on cards. For example, this card might describe how SIC codes and 12 months of ES202 data are arranged on cards. PLOT allows the data to be punched in a variety of ways but requires that the user tell PLOT how he has punched the data. This feature often makes it possible to use data cards which were punched for some other purpose as input to PLOT without repunching them. This can save substantial keypunch expense.

The method of description will be covered in the section entitled "FORMAT Specification" because some aspects of the data must be covered first.

Card 7

This card describes how the second series of data is punched on cards. For example this might explain to PLOT how the BLS 790 series, benchmark 1970, is punched.

Card 8

This card describes how the third series of data is punched on cards. For example, this might explain to PLOT how the BLS 790 series, benchmark 1971, is punched.

Card 9

Columns 1 - 4: DATA
Columns 5 - 80: blank

Cards 10 Through the End of the Deck

The rest of the cards contain the data which the user wishes to plot. The data are grouped by industry: the first group of cards are data for the first industry and will be drawn on the first graph; the second group of cards are data for the second industry and will be drawn on the second graph; etc. Any number of groups of data cards are allowed and so any number of industries may be plotted.

Each group of data cards begins with a 10 character SIC code of the industry whose data are contained in this group. For example, 371bbbbbbb or 371-376bbb or 203-6,9bbb might be the SIC code(s) of the industry (b represents a blank column). Any character may be used but a maximum of only 10 characters may appear. If the SIC code takes less than 10 characters the remaining columns should be left blank. The SIC code may be punched in either of two postions: (1) in the first ten columns of the first data card of the group, or (2) in the first ten columns of a separate card which precedes the data cards. Whether (1) or (2) is done must be indicated by the character punched in column 8 of card 6 as described below.

Next is punched the first data series to be plotted. All months of the data are punched consecutively up to the last month. The number of months punched must be the same as the number punched in columns 8 and 9 on card 5. Any number of columns of the card may be used to punch a data value, but each data value must be punched in the same number of card columns as all of the others. If a data value requires fewer columns than the others, allowance can be made by leaving blank(s) in the columns to the left of the number. If all of the months of data can not be fit on one card continue on to the next card, using as many as is necessary to punch all of the months.

For example, suppose the first series consists of 20 months of ES202 data. Consider what cards should be punched for industry SIC 19. The first 10 columns are for the SIC code but only 2 are needed so columns

3-10 are left blank. All of the employment figures for SIC 19 can be fit into 4 columns per month so starting with column 11, 17 months of data are punched on one card. The final 3 months must be punched on the first 12 columns of the next card.

On the next card starting in column 1 is punched the second data series. The data values may be punched in any number of columns, but they all must be punched in the same number of columns. The number of columns and the arrangement may be different from that of the first series. Any number of cards may be used to punch the second series. Again the number of months of data for the second series must be the same as the number punched in columns 8 and 9 on card 5.

Starting on the next card in column 1 is the third data series. The method is the same as used for the second data series, but the number of columns used for punching a data value may differ from that in the first or second data series.

The rest of the cards in the deck are simply further groups of data cards for plotting further industries. Each group begins with the SIC code of the industry and follows with the data cards for series 1, 2 and 3. Series 1 in each succeeding groups must be punched in the same form as series 1 in the first group; series 2 in each succeeding group must be punched in the same form as series 2 in the first group; similarly for the third series. Any number of data groups may be in the deck.

2.4 FORMAT Specification

Card 6 always begins with 8 characters which tell PLOT how to read the SIC codes in the data groups.

columns 1-7: (2A4,A2

Column 8: either "," or "/" depending upon whether the SIC code in the data group is punched on the same card as the first data series (punch ",") or is punched on a separate data card which precedes the first data series (punch "/"). The other colums of card 6 describe the way in which data series 1 is punched. Each card of data is described with combinations of characters of the form nFm.

n is the number of months of data on that card.

m is the number of card columns used for each data value.

For example, 11F6 means 11 months of data, each requiring 6 columns for a total of 66 columns on the card. If the first data series requires more than one card further combinations of the form nFm can be placed on card 6 to describe them. All such combinations should be separated by a "/". The last character on card 6 should be ")" to indicate the entire first series has been described.

An example of a correct card 6 might be:

(2A4,A2,11F6,4F6)

This card tells PLOT "the SIC code is punched on the first ten characters of the first data card which continues with 11 months of data, each month requiring 6 columns. The next card contains 4 more months of data, each requiring 6 columns." The plots will graph 15 months of data and this number should be on columns 8 and 9 of card 5.

Card 7 describes the second series in the same way as card 6 described the first series, except that SIC codes need not be described. For example card 7 might be simply: (15F4). This card tells PLOT "the second data series is punched on one card with fifteen months of data requiring 4 columns for each month."

2.5 Running PLOT and Examining Tabulations

Turn the deck of cards in at the input window of the computing center where they will be read by the card reader. Pick up the output at the output window and examine it for the following details:

1) Be sure the user signed on properly.
2) Make sure the file was created properly. The output must read FILE "name" HAS BEEN CREATED.
3) Be sure PLOT began running successfully. The output must read EXECUTION BEGINS.
4) Be sure the number of months to be plotted is read correctly. The output must read THE NUMBER OF PLOTTING POINTS SPECIFIED IS n where n is the number of months punched on the CONTROL card (card 5).
5) The tabulations must be correct. The number on the leftmost column should be the industry SIC code. The numbers along the top row of the tabulations are the differences between the first

series and the third series for every month. If the number is positive the third series is greater than the first series; if the number is negative the first series is greater than the third series. The decimal numbers underneath the first row are the differences divided by the first series. They can be considered "percentage differences". Moving down the page, there must be a double row for every industry.

6) The last line on the output must say "n PLOTS GENERATED", where n is the number of industries specified on <u>card 5</u>.

If any of the above details are wrong there is an error in the card deck. It must be corrected and the deck resubmitted before taking the next step. However, if the user does this and the file was successfully created on the first run, it already exists and need not be recreated. Instead substitute for <u>card 3</u>:

```
$EMPTY FILE
```

where FILE is the name of the file.

2.6 <u>Producing the CALCOMP Graphs</u>

Punch a new card deck:

<u>Card 1</u>

```
$SIGNON CCID
```

<u>Card 2</u>

Password

<u>Card 3</u>

```
$RUN *PERMIT PAR=FILE RO
```

where FILE is the name of the file

<u>Card 4</u>

```
$RUN *CCQUEUE PAR=FILE
```

The output from this deck will include a receipt number for a CALCOMP plot. After the CALCOMP plot has been completed (which may take as long as 8 hours) take the receipt to the output window and receive the plots.

2.7 Final Procedures

Writing on the CALCOMP plots, identify clearly the three series by name and dates.

After everything is completed the file which was created should be destroyed so that substantial rental costs will not be incurred. To do this, submit a deck of three cards:

```
$SIGNON CCID
password
$DESTROY FILE
```

APPENDIX G

A MANPOWER INFORMATION SERVICE

Malcolm S. Cohen and Paul Ray

3.0 Introduction

This appendix discusses the issue of how a Manpower Information Service (MIS) might be established. Our recommendation is for establishing manpower information centers, to provide training, information and services, in a few selected regions of the country. This appendix describes the structure of such a center.

The cost of such a center is likely to run about $2 million per year, excluding grants for special studies or projects. However, such a center could effectively meet the information needs for manpower planning, and research and analysis which are not now being met by state data processing centers. State research and analysis staffs could access the service through the utilization of the MIS time-sharing computer system. The cost of setting up such a center is far less than the cost of upgrading state data processing centers to meet the information needs of manpower planning and revenue sharing. Viewed in this manner, the center could improve the effectiveness of labor market information at a lower cost.

3.1 Functional Location of the Manpower Information Service (MIS)

It is recommended that the MIS be responsive not only to the needs of the Employment Service (ES), but also to other groups that it serves. The Director of the MIS should report to an Advisory Committee that sets priorities.[1] The Advisory Committee should consist of: representatives from states that are serviced; one representative from each state and local manpower planning organization; a representative from any other Federal agency contributing major funding; and one or two academic experts. The center should affiliate or cooperate with a university, so that a formal degree program can be established.

The MIS would facilitate the development of labor market information systems, offer training to manpower planners, service requests for manpower information, develop software for use by manpower planners and carry out research related to manpower planning. The center would offer graduate level academic programs in areas such as labor market analysis and manpower planning, seminars and short conferences. Academicians, manpower planners and employment service personnel would be invited to carry out both short and long term research projects.

---

[1] A typical priority question might be: What fraction of resources should go into training?

3.2 Relation of MIS to State Employment Agencies

The relation between the Manpower Information Service and state employment agencies will evolve over time. As new demands are imposed on the Employment Service by manpower revenue sharing, some of these demands can be met by state agencies, while others can best be met by the Manpower Information Service. Each state Research and Analysis Director will not be frustrated by the inability to explain his new information requirements to his data processing chief. Experiments can be undertaken to centralize some information functions which can not now be economically handled by each state. For example, states might collect job applicant data and job orders and transmit the data, by a key to disk entry system like the one used at the Colorado Division of Employment, to a regional data center. The regional data center could then offer on-line job matching to all of the states in the region at a substantial saving over the cost of setting up separate job matching systems in each of the states.

The MIS will also provide training and technical services to state Employment Service personnel, local manpower planners and other Manpower Administration officials at a regional or national level.

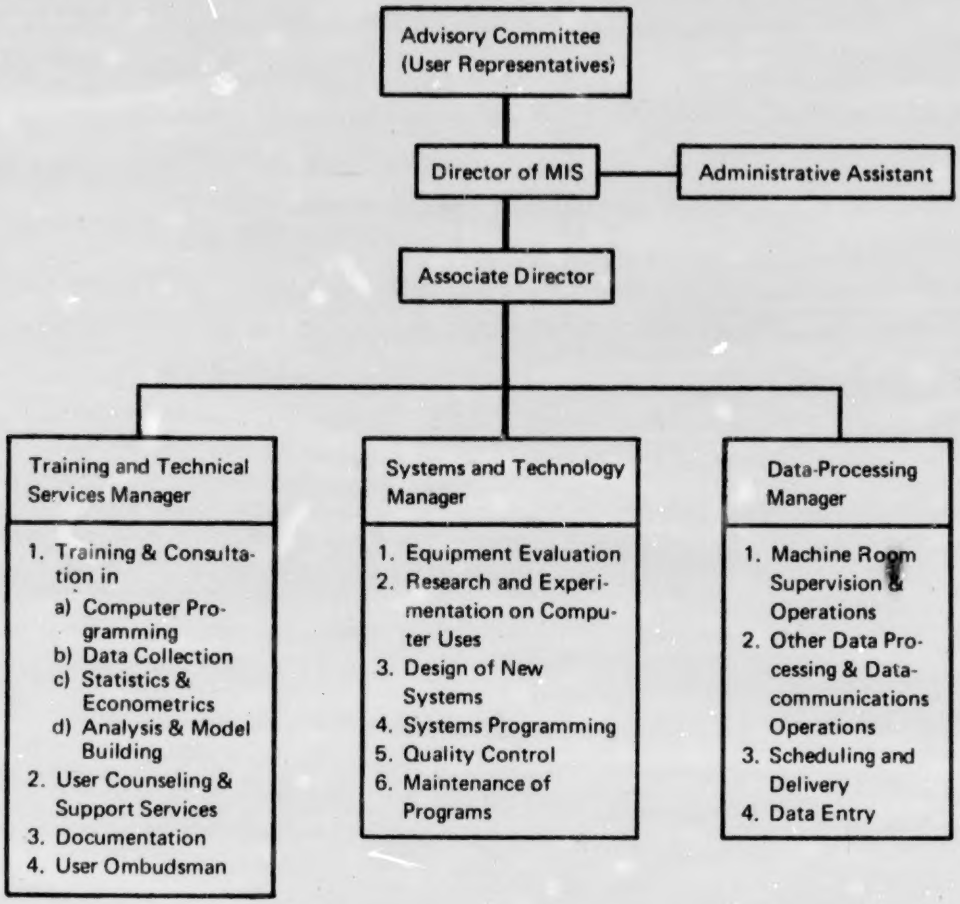
3.3 Internal Structure of the Manpower Information Service

Functional Organization. The accompanying diagram shows the recommended internal functional organization of the MIS. The Director of the MIS reports directly to the Advisory Committee. The relation of the MIS to its clients should be conceived of as providing services, training and consulting arrangements. It is anticipated that many of the personnel of the Training and Technical Services Section would work very closely with the ES Research and Analysis personnel (and equivalent personnel in other agencies) to provide training and consultation services. ES Research and Analysis sections should also be able to expand their professional staff in order to efficiently utilize MIS capabilities and be better able to carry out their own mission with the help of the center. In general, however, the MIS should be conceived of as a relatively "freestanding" organization capable of developing as an information services center according to its own internal logic. (See the proposed organization chart, Figure 6).

It is reccommended that the Director of the MIS have an Associate Director whose responsibilities should include: standing in for the Director; primary responsibility for recruitment of personnel; and supervision of the two "new technology" sections -- the Training and Technical Services Section and the Systems and Technology Section -- and the Data Processing Section. In addition, an Administrative Assistant would be responsible for budget, accounting, supplies, supervision of clerks and secretaries, and so on. The functional responsibilities are divided into three sections.

Training and Technical Services. This section should represent a bridge between new technologies of the MIS and the substantive needs of the users. Its key roles are to serve as a permanent in-house consulting groups and as teachers. The consultation role should be seen as fairly entrepreneurial, i.e., seeking out new areas of applications for the MIS capabilities and working closely with counterpart professionals in those agencies to develop new models, new measures and social indicators, new applications of statistical and sampling techniques and new user applications programs. In this aspect of his role, the Director should encourage advanced research along with other activities, as a way of attracting high quality people. Due to the nature of the tasks, training and consulting should be intertwined. The training aspect of the Section should use many of the same professionals involved in consulting work, plus some personnel from each of the other Sections on a part-time basis. Such training covers a wide variety of activities: a) manpower planning, b) use of the computer, c) new analytic techniques, d) computer programming and e) improving data collection techniques. Their training efforts should include personnel in the state employment service and manpower planners. Such training

Figure 6 Manpower Information Service


Advisory Committee (User Representatives)
Director of MIS
Administrative Assistant
Associate Director
Training and Technical Services Manager
1. Training & Consultation in
a) Computer Programming
b) Data Collection
c) Statistics & Econometrics
d) Analysis & Model Building
2. User Counseling & Support Services
3. Documentation
4. User Ombudsman
Systems and Technology Manager
1. Equipment Evaluation
2. Research and Experimentation on Computer Uses
3. Design of New Systems
4. Systems Programming
5. Quality Control
6. Maintenance of Programs
Data-Processing Manager
1. Machine Room Supervision & Operations
2. Other Data Processing & Datacommunications Operations
3. Scheduling and Delivery
4. Data Entry

ranges from a mechanical "how to do it" for clerks, coders and secretaries, to "why, when and how" for technicians and new professionals. In addition to the activities of training professionals, a certain number of staff should be "circuit riders" who do training in the field, especially in branch offices of the Employment Service, for a substantial part of the time. In addition, a service staff of programming consultants will be needed to explain the use of programs in the system library and the use of various services and options of the computer system. In general, all these personnel should be charged with extending the use of the computer and data bases in cost effective areas, with a responsive servicing policy that draws even non-computer-oriented people into use of the MIS. In addition, this group will be charged with documenting the programs and the uses of the system, plus documenting data access methods and data bases, for all users. This documentation will be important for generalizing and exporting the MIS experience to the other states and to the Federal government. Because the Training Section will have responsibility for dissemination of information to several states, it will be a large unit. The policy of the MIS should be not to require hard evidence of the cost effectiveness of an inexpensive application if the potential benefits are large. This should be especially true in the first MIS center.

Systems and Technology. This section is the core professional computer staff of the new service. It is responsible for evaluating the computer hardware and all technical decisions with respect to it and the operating system. Adaptation of an operating system to the needs of the MIS will be a straightforward task, but one that obviously persupposes familiarity of personnel with both the computer and the operating system. This the staff of this group should be on the scene a year or more before the MIS has its own computer, working toward a smooth transition. They will be systems analysts, designers and programmers who will do systems development and programming, and will control the use of the system. They will also do data base design and maintenance.

Data Processing. This section will primarily be responsible for day to day operation and supervision of the computer. Its responsibilities are machine room supervision and operations, keeping things running smoothly for two shifts. The actual handling of tapes and other storage media, the hourly scheduling of runs, control and use of computer supplies, the handling of data processing and telecommunications equipment, data entry, etc., of a normal large scale computer shop are their responsibility.

3.4 Personnel of the MIS

One crucial aspect of the creation of a Manpower Information Service is the opportunity to introduce in one grand design a whole team of high technology professionals of the kind found in the better universities and in big business. This is a mandatory step in upgrading existing organizations, since adding on new technologies is invariably a disjump in capabilities and in orientations. Incremental solutions almost always meet with internal resistance. In addition, there are "critical mass" problems: esisting organizations find that attracing "new technology" professionals is difficult, because incremental additions of staff would not create the supportive environment needed. The solution adopted elsewhere has invariably been large scale creation of new organizations within the old -- creating a whole system modeled on new organizational concepts. The solution starts by finding a leadership group (the Directors), giving them a mission and an organizational template, or design, plus significant new resources, plus significant freedom to select their own team of professionals (within functional constraints agreed upon by the Director and the Advisory Committee). We are not only suggesting new staff for the MIS, and the rotation of staff within the state employment agencies and other divisions of the Manpower Administration, but we also suggest that the center provide training programs for the Manpower Administration staff.

The Director. It is suggested that the Director of MIS be an experienced administrator with a Ph.D (or equivalent expertise) in such an area as: Computer Science, Economics, Industrial Engineering, Management Science or Operations Research. He should have over five years' experience in government and be fully familiar with manpower problems. A key qualification would be that kind of intellectual/professional leadership that enables him to attract a strong staff of professionals in systems and analysis areas, and enables him to give strong overall direction to their efforts. He must also be able to articulate the MIS's problems and capabilities to administrators and politicians who do not have a strong technical background and, if possible, be a good salesman for the innovative approaches in his mission. It is essential to the success of the concept to MIS that the Director regard his job as being more than merely the Director of a computing center. His mission extends to encouraging wide use of the computer; to the dissemination of manpower information; to the conduct of new research on labor market modeling and social indicators; to the development of new data bases, new measures and new methodologies; and to participating in the upgrading of the skills of the various user groups. His salary should be comparable to a Federal super grade (GS 16-18).

Associate Director. It is suggested that the Associate Director be a strong backup man for the Director, also with a Ph.D. (or equivalent expertise), perhaps with a more specialized technical background, but in the same range of fields as above, and ideally in a different discipline from the Director. He should take primary responsibility for computer management, policy and operations and be the key director of the systems and analysis staffs. He should take major responsibility for recruitment for the organization. Experience in conducting research and development operations, and technical knowledge of computer operations are highly desirable. A knowledge of labor market behavior, while desirable, is not essential. His salary should be in the GS 15 range.

Administrative Assistant. It is suggested that the Administrative Assistant's function is to take the paper work load off the above positions. An M.B.A or M.P.A. is desirable. The salary should be in the GS 13 range.

The Three Sections. Except for Section Managers, the numbers given refer to full-time equivalents (F.T.E.'s) rather than to numbers of personnel. There should be provision for part-time personnel and persons with joint appointments.

Training and Technical Services Section. It is suggested that the professional staff of this section should be regarded as having qualifications and job descriptions surpassing those of most state employees, so that specified qualifications and salaries must be kept competitive regardless of tradition. One way of recruiting such personnel is to encourage two year appointments and joint appointments with universities. The Manager of this Section must be carefully selected by the Director. He must have a Ph.D. and preferably be an economist, econometrician, or an operations researcher. Experience in college teaching and in the labor analysis area is essential, as is ability to direct a high level staff. His salary should be at the GS 15 level. Other personnel are as follows:

Senior Professionals GS 13 to GS 15

F.T.E.	Job
1	Econometrician
1	Labor Economist
1	Industrial Engineer
1	Operations Researcher
1	Statistician with sampling experience

Junior Professionals GS 9 to GS 13

F.T.E.	Job
2	Technical Writers
1	Information Service Librarian
2	Technical Instructors for Programmer & Systems Training[1]
2	Statisticians or Social Scientists for Analysis Training
2	Statisticians or Social Scientists for Data Collection Training
3	Programmers and/or Social Scientists for User Services and Counseling[1]

The Systems and Technology Section. It is suggested that the choice of systems personnel be highly adapted to the demands of the computer configuration. The Manager of this Section must have a strong computer science or related degree, M.S. to Ph.D., from a good university program. He must have knowledge and experience in on-line time-sharing computer systems with virtual memory organization. Experience in directing a systems group, in design of systems architecture and systems software, and in supervision of applications programmers are all essential. His salary should be at the GS 15 level. Other personnel are as follows:

---

[1] Should be joint appointments with other sections.

F.T.E.	Job
2	Systems Analyst
1	Data Communications Systems Planner
1	Systems Designer
2	Programmer Analysts
3	Programmers (Application or Systems)

Data Processing Section. The Data Processing Manager should have a B.S. and preferably an M.S. in Computer Science or a related field, or its equivalent, with a strong background in hardware operations on the chosen computer configuration, and with a special competence in data communications. His salary should be in the GS 14 range. Other personnel (assuming two shifts of operations) are as follows:

F.T.E.	Job
2	Lead Operators
4	Machine Operators
4	Input Clerks for keying data to disk, tape, etc.
2	Library Functions Operators

Clerical Staff. It is assumed that a desirable secretary-clerk staff to service such an information service will be as follows:

F.T.E.	Job
3	Secretaries for the two Directors & Administrative Assistant
1	Secretary for each Section Manager
4	Secretary/Clerks in a pool

3.5 Overall Budget Parameters

It must be noted that evaluating the cost-effectiveness of this proposal necessitates comparing new levels of output being produced with alternative approaches for obtaining the same information and services.

It is anticipated that equipment cost would be about $1 million per year and personnel costs are estimated to also be $1 million per year. These estimations will vary by inflation and exact staff configuration and level of activity.

New facility costs and office equipment costs could range from as little as an additional $100,000 to $1,000,000 depending upon ability to convert existing space, or lease new space, etc.

3.6 Conclusion

A Manpower Information Service would provide valuable aid to labor market information users through its capacity to provide new services and training. Only through a reasonable investment in human and other resources can adequate manpower information be provided.

APPENDIX H

A SET THEORETIC DATA STRUCTURE AND RETRIEVAL LANGUAGE

William R. Hershey and Carol H. Easthope*

4.0 Introduction

The development of a data structure for use with labor market information has led to a system which is capable of handling many types of data bases. The Labor Market Information System (LMIS) Project is being funded by the U.S. Manpower Administration to study the feasibility of implementing a nation-wide information system for the storage and retrieval of labor market data and to build a labor market model. The characteristics of a good labor market information system are not unlike the characteristics of data bases in many other applications. Therefore many of the features discussed here should be of widespread interest.

The goal of an accurate model of the labor market almost demands an efficient and accurate information system to supply the model's input data. A second purpose for developing an information system is to provide an extremely simple retrieval system to be used by the state Employment Services for the access of data that is not available through their "canned" report generating programs.[1]

Two of the primary criteria that our information system must satisfy are generality and compatibility to handle existing data files for different surveys. An interactive system is highly desirable, since it allows our economists to ask questions of the data bases quickly, and since we can support and maintain our system on one computer which can be called easily by the states that are participating in our project. The Michigan Terminal System (MTS) on which the program is run is also a very good interactive environment.

---

[*] The Labor Market Information Systems Project is sponsored by the Manpower Administration, Office of Research and Development, United States Department of Labor, under contract no. 71-24-70-02. The views represented in this paper are the sole responsibility of the authors and do not necessarily reflect the views of the Department of Labor.

[1] We are presently supporting our systems for three states: Michigan, Wisconsin and Colorado.

The uses to be made of our retrieval program require the utmost simplicity at the user level. Hence, the instructions are entered as English sentences rather than in fixed formats. There is an option to correct spelling mistakes, and the program is self-documenting via commands from the user.

4.1 General Organization and Function of the Program

The traditional way to implement a retrieval system has been to determine what questions are to be asked of the data and then to write a program and data structure to answer those questions. Thereafter one is locked into that class of questions. However, in our applications, and indeed in many other applications, it is not known what questions will be posed. Nor is the data in any particular format. We had to look into new ways of designing our system, so that any question could be posed with any data.

We have available on the University of Michigan computer a generalized retrieval program called Set Theoretic Data Structure (STDS), developed by Set-Theoretic Information Systems Corp., Ann Arbor, Michigan. This program is composed of a number of very efficient routines which treat the data bases as sets and perform set operations on them, e.g., union, intersection, restriction, etc. STIS Corporation holds the view that there exists an information enviornment to which questions can be directed, a machine environment in which the data resides, and that "Any data structure is actually an isomorphism between a machine environment and an information environment preserving the functional aspects of each" [3]. They feel, however, that the usual data structures do not preserve the functional differences; the information environment is made to look like the machine environment.

The problem is to find a data structure that can map the myriad relationships of the information environment into the algorithmic, procedural world of the machine. Such a structure could be a set-theoretical model. STIS has extensively investigated the proposition the general information requests can be abstracted to set operations. For example, if an information request is expressed as a set operation, $\circledast$, given data sets A and B, with the retrieved result as set C, then the abstraction can be stated:

$$A \circledast B = C$$

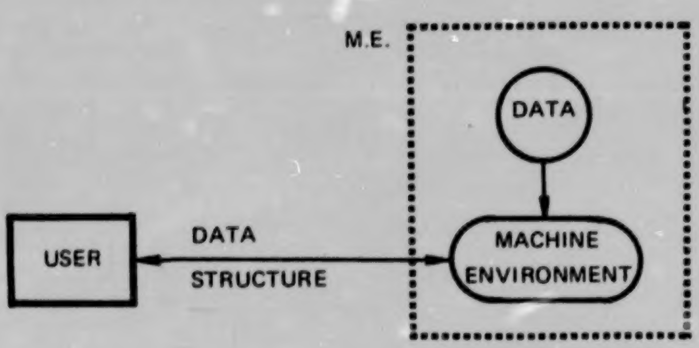
To define this abstraction and thereby prove that $\circledast$ is a valid set-theoretic operation, it is further stated that an element x is a member of C if and only if there is a truth function relating A, B, and x, i.e.,

$$C = (x \mid \psi(A,B,x))$$

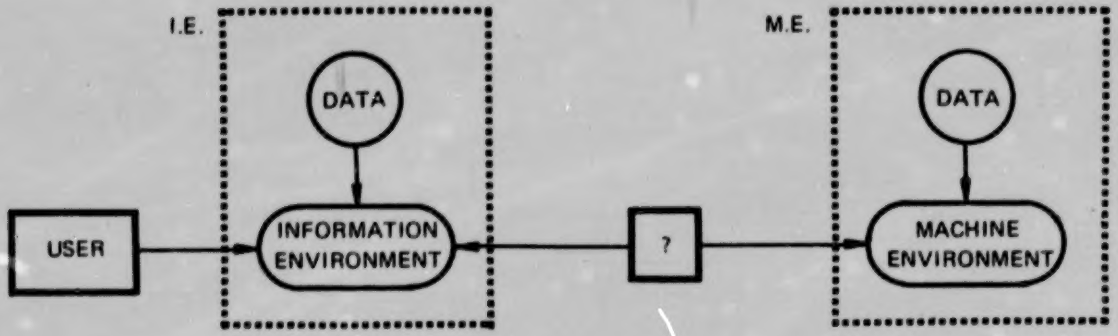
It is the fact that the function is decidable that makes $\circledast$ a set theoretic operation and assures that C is defined. Any algorithm or procedure that decides ψ is valid. Therefore any convenient and/or economic machine representation of data can be mapped into the set-theoretic information environment. (See Figure 7) hence the STDS routines are essentially machine independent, although the initial versions require the paging facility for virtual memory with the MTS operating system on the IBM 360/67.

Our retrieval system, which we have named MICRO, calls on the STDS routines to do set-theoretic retrieval operations on the data. Essentially then, our program is an interface between the user and the STDS routines which do tha actual manipulation of the data (see Figure 8). This interface includes all the facilities needed to make data retrieval a simple task for users who are totally unfamiliar with computers. There is a syntax analyzer to parse the input commands, a dictionary to associate fields and records in the data with words that the user understands, 1/0 and file manipulating routines that prepare the data in the form needed by STDS, error diagnostics and even automatic error correction, self-documentation to aid the confused or forgetful user, and facilities for preparing the output for the terminal, for peripheral storage, or in a format suitable for input to statistical routines. It is with this MICRO interface program and its affiliated data structures that this paper concerns itself; literature describing the STDS routines is available from STIS Corp. [14]. (See also [4] and [5].) In our experience thus far it has been shown that the user interface is an extremely important part of the total retrieval package, since with a more difficult program the unsophisticated user would be helpless and not likely to utilize the data most effectively.

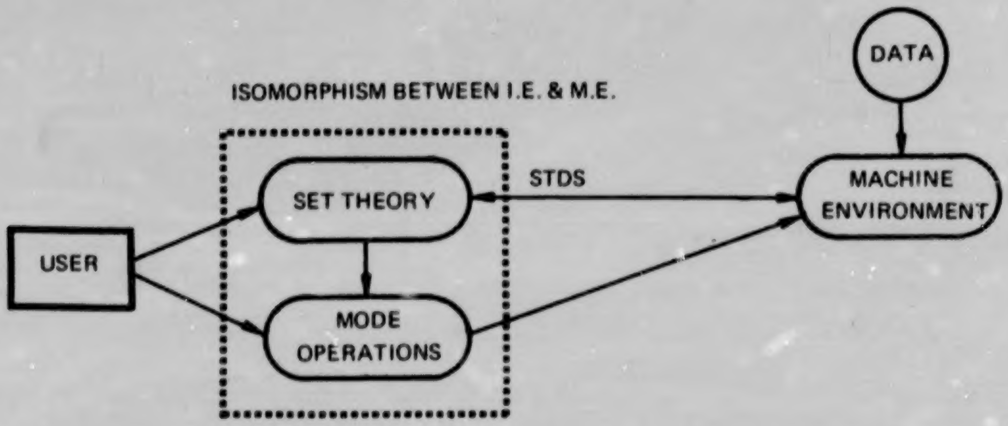
Figure 7 Contrasting Data Structures


M.E.
DATA
USER
DATA
STRUCTURE
MACHINE
ENVIRONMENT

USUAL DATA STRUCTURE - User can only make retrieval requests as structured by the machine environment.

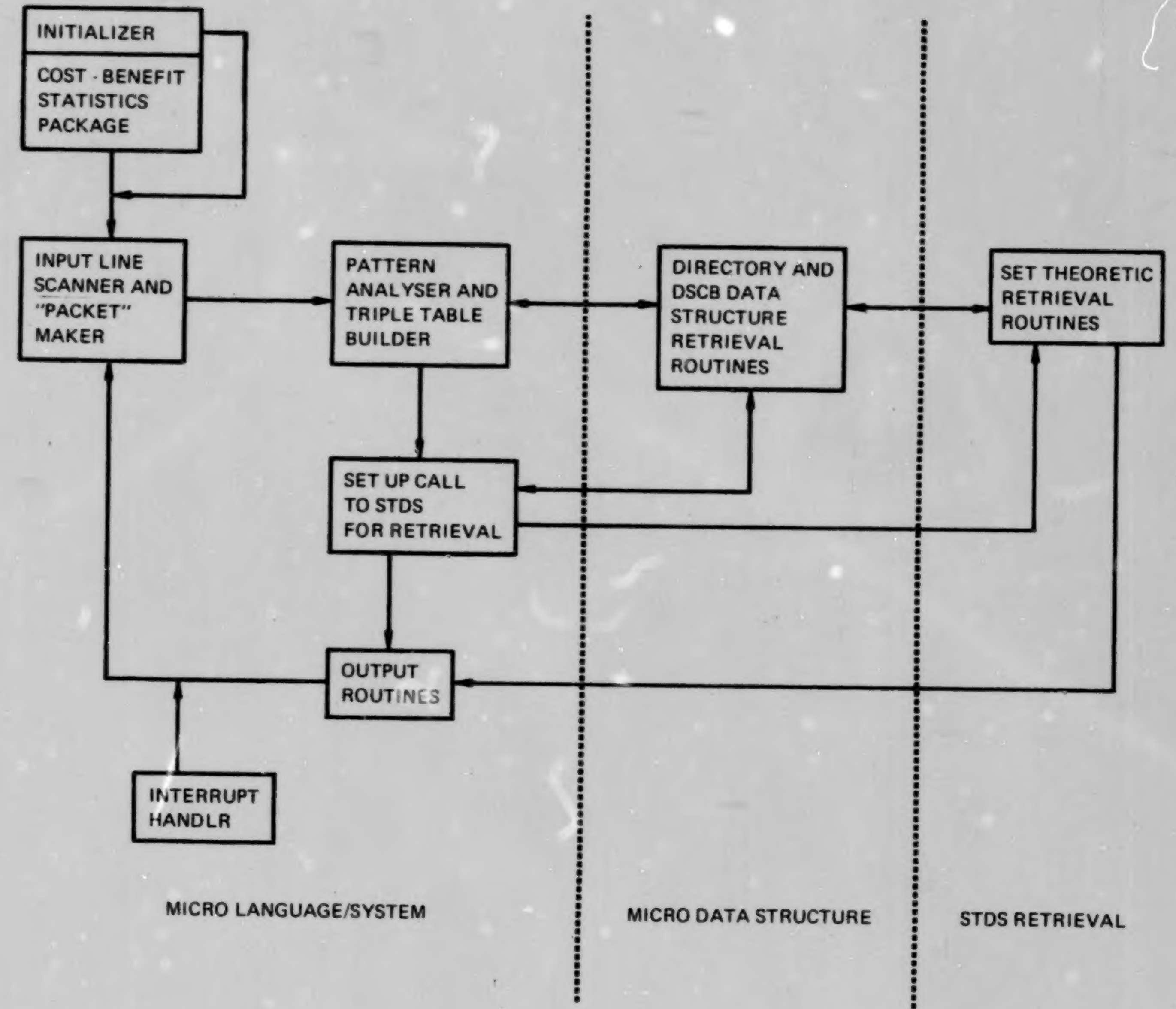

I.E.
M.E.
DATA
DATA
USER
INFORMATION
ENVIRONMENT
?
MACHINE
ENVIRONMENT

IDEAL DATA STRUCTURE - User can make any request, limited only by and inherent limitations in the relationships within the data itself. There is the need for an isomorphism between I.E. and M.E.


DATA
ISOMORPHISM BETWEEN I.E. & M.E.
SET THEORY
STDS
MACHINE
ENVIRONMENT
USER
MODE
OPERATIONS

SET THEORETIC DATA STRUCTURE - User expresses request as set operations, which give him full retrieval power. The STDS and mode operations isolate the user from M.E.

Figure 8 MICRO Interface To Set Theoretic Operations


INITIALIZER
COST - BENEFIT STATISTICS PACKAGE
INPUT LINE SCANNER AND "PACKET" MAKER
PATTERN ANALYSER AND TRIPLE TABLE BUILDER
DIRECTORY AND DSCB DATA STRUCTURE RETRIEVAL ROUTINES
SET THEORETIC RETRIEVAL ROUTINES
SET UP CALL TO STDS FOR RETRIEVAL
OUTPUT ROUTINES
INTERRUPT HANDLR
MICRO LANGUAGE/SYSTEM
MICRO DATA STRUCTURE
STDS RETRIEVAL

4.2 Structure and Content of the Data Bases

The STDS routines work equally efficiently on any data. Therefore, retrieval can be done in any mode. But since our data bases tend to be large (up to one million bytes), we convert data from characters to binary numerical form before using it with the program. While the cost of the one-time conversion is sometimes high, the conversion effects considerable savings in virtual memory and execution time charges for the retrieval program.

The MICRO representation of the data can be visualized as a matrix in which the rows represent different records or "cases", and the columns are fields in which are recorded characteristics for each record; however, the actual data representation is in set-theoretic form. In labor market applications the rows typically represent job applicants or employees and columns (called Fields) designate fields such as age, sex, and income or type of industry, number of employees, and payroll. Each column, i.e., Field, can have certain "Categories" which stand for different values of the specified Field. For example, the Field SEX would have the categories MALE and FEMALE.

The Categories are coded in the data structure as numerical values, and they are translated to words that the user understands at retrieval time via a "dictionary". Field names are also included in the dictionary, since STDS does not currently keep track of these names. The retrieval is one by using the byte position and field length of the Field in the record; hence each Field name in the dictionary has associated with it the length of the Field (one to four bytes) and the Field's byte position within the record.

Some of the information in a data-base may be meaningful as numerical information, i.e., not translated into a Category word. Examples would be age or income. In these cases no Category recodings are provided, and MICRO converts the binary number to decimal for printout when necessary.

4.3 Use of the Program and Data-Bases

The set-theoretic approach allows great flexibility in the retrieval operations. Four major retrieval operations will be discussed briefly here. The MICRO command keywords for these operations are FIND, XTAB, SELECT, and RESTRICT. All of these operations result in the formation of a RESULT set, i.e., a subset of the original set. This RESULT set can be renamed for further manipulation or storage by use of the NAME command.

The FIND command extracts a subset of the data-base by matching specified Category information under a specified Field or Fields. Hence the resulting subset consists of selected rows of the data matrix. Logical combinations of Fields are possible, as shown in the following examples with a data-set called "SOC-SEC".

```
FIND IN SOC-SEC WHERE SEX IS FEMALE.
FIND IN SOC-SEC WHERE SEX IS MALE AND RACE IS WHITE.
FIND IN SOC-SEC WHERE AGE IS BETWEEN 20 AND 30.
FIND IN SOC-SEC WHERE OCCUPATION IS CARPENTER OR PAINTER OR
   PLASTERER.
```

After finding the RESULT subset, MICRO stores the subset temporarily and prints out a count of the selected cases, i.e., the number of records whoe Category values under the specified Fields conform to the logical combination designated in the FIND command. The RESULT subset has the same record length as the original set, i.e., all Fields in the original set are included.

The XTAB command can do a cross tabulation of frequence of co-occurrence of codes under specified Fields. The number of Fields XTABed can range from two to six, and with this command the entire RESULT set is printed out. The following example will illustrate the command and a typical resulting printout.

```
XTAB IN SOC-SEC SEX BY RACE.
```

Printout:

```
SEX            RACE           COUNT
MALE           WHITE         25,235
MALE           NEGRO          5,243
MALE           OTHER            451
FEMALE         WHITE         27,457
FEMALE         NEGRO          6,105
FEMALE         OTHER            347
```

The SELECT command selects a subset by picking out specified Fields in all records. In other words, the subset consists of designated columns of the data matrix. Note that the logical OR operation is illegal in SELECTing Headers.

SELECT IN SOC-SEC SEX AND RACE AND OCCUPATION.

SELECT IN EMPLOYERS PAYROLL AND INDUSTRY.

The RESTRICT command is usually used to match records between two different data-sets. Both data-sets will have Fields with ID numbers that match for corresponding records. The program finds all records in a data-set whose values under a specified Field match any of the values under a specified Field in another data-set. For example,

RESTRICT IN SOC-SEC WHERE IDNO IS ID IN JOB-APPS.

In this example SOC-SEC is a data-set containing records of all individuals residing in a given area, and IDNO is a Field name for a unique identification number for each individual. JOB-APPS is a mythical file of job applicants at a State Employment Service, and ID is the Field for the unique identification number for each applicant. With the command shown one could find records in the SOC-SEC data-set for only those individuals who are in the JOB-APPS data-set. The cardinality of the RESULT set is printed out, i.e., the number of records in SOC-SEC whose values for IDNO match values of ID in the data-set of JOB-APPS. The length of records in the RESULT set will be the same as the records in SOC-SEC, the RESTRICTed data-set.

Other MICRO commands allow the user to SAVE a RESULT set permanently for future reference, to PRINT all or part of a permanent or temporary data-set, to DESTROY data-sets that will no longer be used, to DELETE data-sets temporarily to save on the cost of core storage during operations on other data-sets, and to ask for documentation about the contents of data-sets as well as the MICRO commands themselves.

4.4 Program Control and Use of the Data Structure

Having looked at MICRO's general structure and operation, we can proceed to study the data structure itself in more detail. The data structure is divided into three parts: the directory, the dictionary, and the data itself. We will treat each separately. All three components are stored on disk all the time, and they are brought into core separately when needed. There is no difference between the representation on disk versus that in core, except that the disk addresses (note pointers) are replaced by core addresses when the information is read into core.[2]

4.5 Directory

The directory is simply a list of available MICRO data-sets. It contains the names of data-sets and the location of their dictionaries (see Figures 9 and 10). It also informs MICRO whether the data-set can be destroyed. Typically, several individuals have access to the same data-sets, and not to those of another group. There is also a master directory which lists datasets that anyone may use. These master data-sets either are for demonstration purposes, or contain non-confidential information that is to be shared among different user groups.

When MICRO begins execution, it reads the user's directory into core and then the master directory. It combines the two into a simple forward ring (see Figure 11), creating a "universe" to which the user has access. Whenever a temporary data-set is created with MICRO, a directory entry for it is inserted as the second element of the ring. The premise is that a temporary data-set is more likely to be referenced again and therefore should be near the beginning of the ring.

A directory entry for a temporary data-set is not written to the disk directory list unless a SAVE command is issued. In that case it is simply written as the last record of the user's disk directory. Therefore, there is no correspondence in order between a disk directory and the in-core directory.

When a temporary data-set is DELETEd, the entry is removed from the ring. (A permanent data-set only has its data deleted from core, not its directory entry.) Hence there are never any directory entries for temporary data-sets on disk. However, when a permanent data-set is

---

[2] By core we mean really virtual memory. Our computing system has a paging facility to continuously copy core to drum and vice versa according to user demand for core. The combination of core and drum storage is called virtual memory.

Figure 9 Directory Entry Format On Disk

Data-Set Name
N-Tuple Name
File Name (on disk) for Dictionary
Signon ID where dictionary resides
for future use
Note Pointer to DSCB
for future use
for future use
Protection Switch

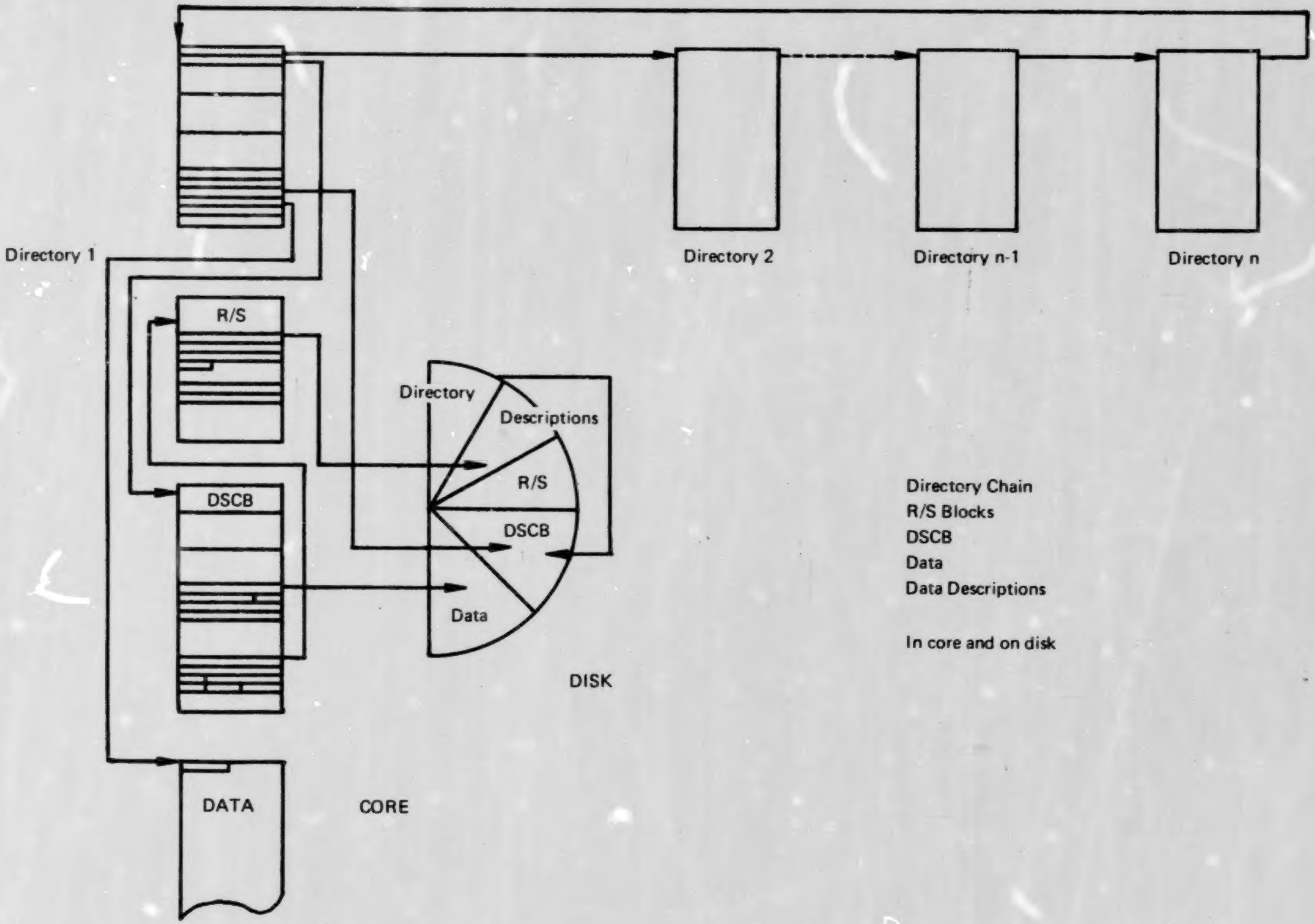
(One entry for each data-set)

Figure 10 Directory Format In Core

Pointer to next directory entry
Pointer to DSCB if DSCB is in core
Data-Set Name
N-Tuple Name
File Name (on disk) for Dictionary
Signon ID where dictionary resides
FDUB for the file, if opened
Note Pointer to DSCB
Pointer to Data-Set in core, if there
for future use
Protection Switch

(One area for each data-set)

Figure 11 MICRO Combines Dictionary And Directory


Directory 1
Directory 2
Directory n-1
Directory n
R/S
DSCB
DATA
CORE
Directory
Descriptions
R/S
DSCB
Data
DISK
Directory Chain
R/S Blocks
DSCB
Data
Data Descriptions
In core and on disk

DESTROYed (a DESTROY of a temporary defaults to a DELETE), the disk directory entry is removed, along with the in-core directory area, the dictionary, and the data.

4.6 The Dictionary

The dictionary itself has three parts: the data set control block (DSCB), the R Blocks and S Blocks, and the data descriptions (English language descriptions of the data-set and its Fields and Categories). (See Figures 12, 13 and 14.) There are only slight restrictions on the order of the dictionary components. They are usually arranged with all data descriptions first, followed by R and S Blocks, and then the DSCB and tape mounting information (if the data is on tape rather than disk). The directory entry for a data-set dictionary contains a note pointer (disk address) to the location of the DSCB record in the file where the dictionary resides. The DSCB in turn has note pointers to the R/S Blocks, and R/S Blocks have note pointers to the data descriptions.

MICRO reads in a DSCB and it associated R/S Blocks on demand, either explicitly when a USE or ACQUIRE command is issued, or implicitly when a retrieval command such as FIND is issued. Once in core, the DSCB remains there for the rest of the session, unless a DESTROY or SAVE command is issued.[3] The directory area for the data-set has a pointer which is set to point to the DSCB in core (see Figure 11). The note pointers in the DSCB are also changed to core addresses for its R/S Blocks. However, the data descriptions are never in core permanently. The note pointers in the R/S Blocks remain as disk addresses all the time. MICRO, in response to a DESCRIBE... Command, does a direct-access read from disk with the appropriate note pointer.

When the structure and format of the directory and dictionary were first designed, it was felt that room should be left for after-thoughts and future expansion. That is the reason for the abundance of "spare"

---

[3] The DESTROY command releases core for the DSCB and data-set, as well as destroying the dictionary and data-set files. The SAVE command, while creating dictionary and data-set files, releases the current core for DSCB, R/S Blocks, and data-set. The SAVEd data-set's DSCB is read in again on demand if it is again referenced.

Figure 12 Data Set Control Block (DSCB)

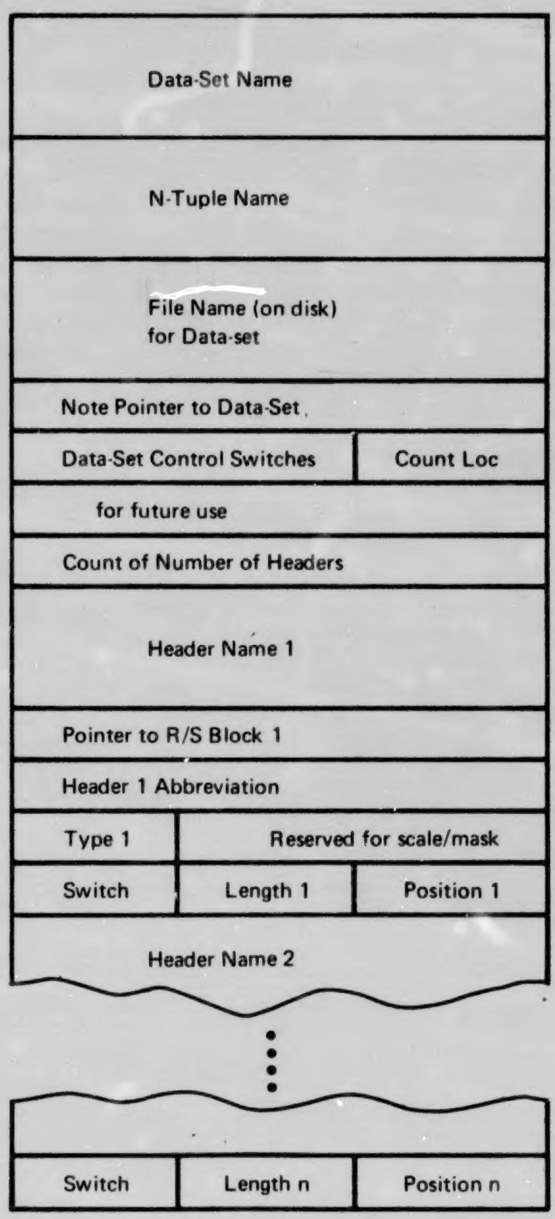
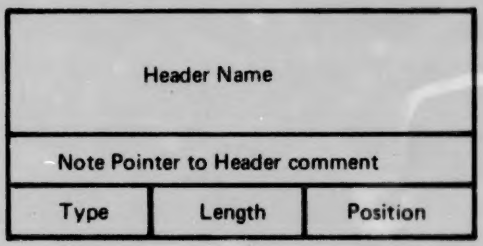
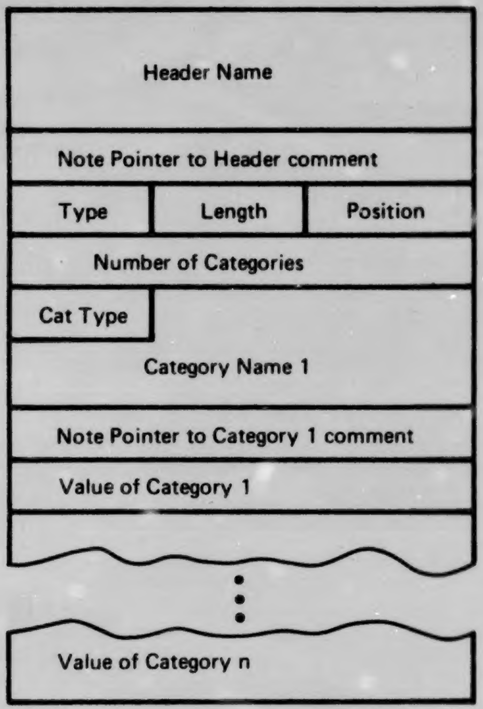

Data-Set Name
N-Tuple Name
File Name (on disk)
for Data-set
Note Pointer to Data-Set
Data-Set Control Switches
Count Loc
for future use
Count of Number of Headers
Header Name 1
Pointer to R/S Block 1
Header 1 Abbreviation
Type 1
Reserved for scale/mask
Switch
Length 1
Position 1
Header Name 2
Switch
Length n
Position n

Figure 13 The R Block


Header Name
Note Pointer to Header comment
Type
Length
Position

(For Headers with numerical information only)

Figure 14 The S Block


Header Name
Note Pointer to Header comment
Type
Length
Position
Number of Categories
Cat Type
Category Name 1
Note Pointer to Category 1 comment
Value of Category 1
Value of Category n

(For Headers with coded Category information) bits and extra spaces in the various control blocks. In fact, since the first demonstration version (late March, 1971), about one third of the original "spare parts" have been utilized. At the same time, data-sets created at the beginning of the project are still compatible with new and improved versions of MICRO.

4.7 The Data-Set

The structure of the MICRO is extremely simple. It can exist on disk or tape. The first record is a specially encoded fullword that specifies the data-set's n-tupple length (record size) and its cardinality (number of records). The data immediately follows in 32K-byte records (rows of the data matrix are stored contiguously).

A data-set record can be a mixture of binary and character fields, where the field length can be from one to four bytes. The field length is a limitation of the current STDS routines. (A more recent version of STDS will allow a length from one bit to 32767 bytes.) A great deal of effort is put into the analysis stage of preparing the data as a set form with an interactive version of STDS.

The data-set is not read into core until needed for a retrieval. Once in core, the data-set will remain there until it is DELETEd or DESTROYed, (or in the case of temporaries, SAVEd.) If MICRO needs more core than is available, the program will, if possible, release the core of permanent data-sets (they can always be read in again), and use that. If there is not enough core to release, MICRO will ask the user to DELETE or SAVE some temporary data-sets.

In core the data-set is one contiguous block containing a double word specifying the original size request for the data and how much space is not being used, then the fullword specifying the length and cardinality of the data-set, and then the data.

Data-sets that are confidential can be created and stored in "scrambled" form. This process involves the use of a password which is used in scrambling the data. A dump of the datafile would show only meaningless numbers. MICRO will request the password from the user when it reads in the data, and the data will be unscrambled at the time it enters core. If an incorrect password is given (the user has only one chance), retrieval requests proceed with garbage results.

Presently the size of a data-set cannot exceed 1,044,480 bytes because of STDS program restrictions and MTS operating limitations. But a newer version of the STDS routines will allow buffered retrieval operations for larger data-sets. The data-set can exist on the same disk file as the dictionary or on a different file. Usually it is the latter case, since this allows more efficient use of disk space.

Temporary data-sets that are created as the result of a retrieval operation exist as regular MICRO data-sets with their own directories, DSCB's, and R and S Blocks. Much of the information in these structures is the same as that of the parent data-set. Whenever a temporary set is SAVEd as a permanent set, this information is used to generate the disk equivalents for directory, DSCB, R/S Blocks, and data descriptions. A temporary data-set which is a descendent of a scrambled data-set will be saved as a scrambled set, i.e., a password is requested to be used in encoding the output set.

4.8 Limitations and Plans for Improvement

One of the limitations of the MICRO system is the inability to handle arithmetic operations on numerical data values. While this capability is desired for a future version of the program, it is possible now to achieve the same results by using the MICRO command, WRITE FOR ANALYSIS... This command prepares the RESULT data-set in a form suitable for input to a statistical program on the U. of M. computer called CONSTAT. The desired operations can then be specified through CONSTAT.

A second limitation is in the editing and updating capabilities of the program. But we have immediate plans to implement a COMBINE command (union in set theoretic terms). This command will allow updating, modification, and merging of data-sets. Another modification further in the future will provide even greater editing capabilities.

Data manipulation on the bit level, an improved syntax analyzer, and facility with especially large data-sets are other desired improvements.

The program is designed so that routines for data structure control, I/O, space allocation, and retrieval of elements from the directory, DSCB, and R/S Blocks are separate from the routines for the syntax analyzer, printing, and retrieval of data-sets. By keeping the two major parts of the system separate, we are more easily able to continually upgrade and improve the system.

APPENDIX I

MICRO INFORMATION RETRIEVAL SYSTEM
(VERSION 3.9)
TECHNICAL REFERENCE MANUAL

Michael A. Kahn
Donald L. Rumelhart
Boyd L. Bronson

5.0 SYSTEM BASICS

5.0.1 Introduction

MICRO is an interactive information retrieval system, designed to be used on remote terminals ("typewriter" consoles). The system is interactive in that the user issues a command request to the system and waits for the system to respond before issuing another command. This is particularly convenient since succeeding commands are frequently dependent upon the results of the previous queries.

The MICRO Information Retrieval System has general applicability to a number of problems. It could be used for job matching, medical research, inventory control or retrieval of management information. The system is very powerful in terms of the complexity of the requests which can be made. Also the structure of the commands is English-like which makes the system easy to learn and easy to remember.

The system is limited, however, in that it must be run on an IBM 360/67 using MTS (Michigan Terminal System). This system is resident on computers at the University of Michigan and at Wayne State University. Both of these computer installations are accessible via regular telephone lines.

5.0.2 Concepts and Facilities

Data Sets and Dictionaries

In MICRO different data collections are called data sets and the collection of data sets is referred to as a data base. A data set is simply a collection of records. Each record is composed of fields. MICRO can have many data sets available to it. Each data set is referenced by a name and is stored on a disk or tape file (auxiliary storage device). A user only needs the data set name to reference a data set. A list of the available data sets can be obtained at any time during a MICRO session.

MICRO data sets are self-describing; that is, descriptions of the contents, formats and names of the data items are themselves a part of the data set. These descriptions are called the dictionary. Information about the names of fields and about various attributes of the data set(s) is stored in the dictionary.

Therefore, a MICRO data set consists of records of data and a dictionary. The collection of records alone is called an STDS (Set-Theoretic Data Structure) set, and is in a form to be used by the STDS system.[1]

Each field within a record is given a name, called the "field name." Fields can be thought of as the attributes of each record. A field name is a string of no more than sixteen characters. For example, a field of social security numbers could be given the field name "SOCSECNUM". Since long field names can be cumbersome, an up-to-four-character abbreviation is allowed. The abbreviation for the above example might be "SSN".

The contents of a field in MICRO are called the values of the field. For example, the value of the field "SOCSECNUM" for the first record may be 123456789. However, certain kinds of data can best be thought of as falling into categories. Therefore, some values of fields may be referenced by category names. For instance, a field called SEX may have actual values of 2 or 3. However, it would be very convenient to refer to these values with the names MALE or FEMALE. Thus MICRO has a name-to-value association in the dictionary and data may be referred to by actual values or by category names, if any exist. Further, MICRO will use the category name when printing whenever possible. Category names may be up to eleven characters in length.

Data sets may or may not possess one or more of the following properties:

(1) Destroyable - certain data sets can never be destroyed while others can be permanently destroyed.

(2) Replaceable - certain data sets can never be replaced while others can.

(3) Scrambled - a confidential data set may be maintained in a specially encoded ("scrambled") form. If a data set is scrambled, a special password must be entered to

---

[1] STDS is a program product of Set Theoretic Information Systems Corporation, 117 N. First Street, Ann Arbor, Michigan 48108.

gain access to that data set.

(4) Permanent/temporary - a permanent data set is one that is available from one MICRO session to another. A temporary data set is one that is established during a MICRO session. Temporary sets are not available from one MICRO session to another. A temporary data set can be made permanent by using a SAVE command.

(5) Cross Tabulated - In general, a data set will have an instance of every record even if some records are identical to other records. Occasionally it is advantageous to combine identical records into a single record with a count of the number of such occurrences. This can be done with a CROSSTABULATE command and will produce what is referred to as a cross tabulated ("cross tabbed") set. A cross tabulated set can result in a considerable savings of space. However, it has the disadvantage of treating items in the aggregate which eliminates further use of some of the MICRO commands.

These properties of a data set are assigned when the data set is first created or SAVEd. The scrambled quality can be inherited. If a new data set is created by taking a subset of a scrambled data set, then the new set will also be scrambled.

At any time during a MICRO session there are a number of data sets available. Those data sets are called the universe of data sets. When a user renames a RESULT set (see below) that set temporarily becomes a member of the universe. A SAVE command can be used to add a set permanently to the universe.

Subsets and the Result Set

It is possible to extract from a data set a subset of records meeting certain criteria. This subset is itself a MICRO data set and is automatically given the data set name RESULT. The contents of this set are available until another command which creates a RESULT set is issued. If the user wants to keep a RESULT set, it can be given a new name. This newly named set then becomes a temporary set and is available for the remainder of the MICRO session. If desired, the set can be SAVEd and hence would become a permanent set.

Documentation Facility

A documentation facility (DOC) exists to allow the user's queries to be copied ("echoed") into certain files when the facility is enabled. This enables supervisory personnel to discover which commands are used most frequently and also allows them to analyze the user/computer interaction when the user encounters problems with the system. At present the documentation facility defaults to OFF.

Macro Subsystem Within MICRO

A macro subsystem exists for the purpose of extending the MICRO language and to facilitate the reference of often-used sequences of MICRO commands. It can be accessed directly from MICRO. See the third part of this manual for the details of the macro facility. Descriptions of existing MACROS may be obtained by writing the authors in care of the Institute.

5.0.3 General Information

Prefix Characters

Whenever MICRO communicates with the user a special prefix character is printed as the first character of each line. These are:

(1) minus sign (-): request for a (first) line of a MICRO command.

(2) plus sign (+): request for additional lines of a command.

(3) asterisk (*): MICRO has generated the printed line.

(4) equal sign (=): indicates that the macro subsystem has generated the printed line.

Format of Commands

A MICRO command can be any number of lines in length. A period (.) is used to indicate the end of a command. A command consists of:

(a) the command name,

(b) a series of words such as data set names, field names, and special keywords as is indicated in the prototype of the command description. Each word must be separated by one or more blanks,

(c) a period.

Notation Used in This Manual

The following conventions are used in this manual:

(1) Any word that appears in all capital letters must appear (as is) in the command.

(2) Words or phrases that appear in angle brackets (<>) should be replaced by an actual value or the appropriate name. For example: <field name> might be replaced with SEX or RACE.

(3) Words or phrases that appear in square brackets ([ ]) are optional and may be omitted. For example in the CALL command, [PARAMETER=<arg 1>,...,<arg n>] may be omitted if the subroutine requires no additional parameters.

(4) Words separated by a vertical bar (|) indicates that only one of the terms can validly appear in the command. For example, AVE|TOT indicates that either AVE or TOT should be used, but not both.

(5) The ellipsis (...) is used to indicate that a word-type or phrase-type may be repeated as many times as is required.

Further these conventions can be used in combination. For example, [AVE|TOT] indicates that either AVE, TOT or neither would result in a valid command.

Abbreviations and Synonyms Acceptable in MICRO

The following symbols may be used interchangeably in a MICRO command:

(1) DATA SET[S] | DATASET[S] | SET[S]

(2) CATEGORY | CATEGORIES

(3) FIELD | FIELDS

(4) AND | &

(5) COMMAND | COMMANDS

(6) ALSO WHERE | OR WHERE | ;

(7) OR | |

In addition, most MICRO commands have synonyms. The synonyms for a command appear with that command's description.

5.0.4 Security Features of MICRO

There are several ways by which confidentiality is maintained. As was discussed previously, some data sets are scrambled and require a password to be accessed. Data sets are available only to users who know the proper index (password) required to access directory information for those data sets. (See the GET command.)

In addition to these measures some versions of MICRO require an additional password to use the system at all. If a protected version is being used, the MICRO system will prompt the user for information about the type of terminal being used and for the password. (See the Protection Key Facility.)

Levels of Security

There are five ways in which confidentiality is maintained:

(1) Knowledge of a system (MTS) sign-on ID and appropriate password is required in order to access the Michigan Terminal System.

(2) The user must also know which command will initiate the proper version of the MICRO Information Retrieval System for the intended application.

(3) Certain versions of MICRO require the use of an additional password in order to further use MICRO. (See Protection Key Facility.)

(4) Data sets are available only to those users who know the proper index (password) required to access directory information for these data sets (See the GET command).

(5) Finally, for the most confidential information, a data set can be scrambled (specially encoded according to a key) which requires an additional password for access. It should be noted, however, that this is a very costly protection facility.

Protection Key Facility

A protection key facility (PK) exists to limit access to the MICRO System to authorized users. If PK is ON when the user runs the MICRO system he will automatically be prompted for the type of terminal being used. There are two possibilities: (a) teletype compatible and (b) not teletype compatible. Authorized users will be informed of the distinction. After indicating terminal type, he is prompted for an additional password. If he does not enter the proper password he is disconnected from the MICRO System. This additional password will be given only to those authorized to have access to MICRO and the data sets available. The Protection Key Facility prevents the use of the MICRO System from an unauthorized signon ID and its use from an authorized signon ID unless the proper protection key is used.

5.1 COMMAND DESCRIPTIONS

CALL

COMMAND DESCRIPTION

PURPOSE: To execute a user or system subroutine while in the command mode of the MICRO Retrieval Language and to return to command mode upon completion of the subroutine.

PROTOTYPES AND DESCRIPTIONS:

(1)
```
CALL <subroutine name> [USING <data set name>]
[LIBRARY=<library name>] [KEEP=YES|NO]
[PARAMETER=<arg 1>,...,<arg n>].
```

If USING <data set name> is omitted, then the RESULT data set is assumed.

If LIBRARY=<library name> is omitted, then the MICRO system library is used.

KEEP=YES|NO is used to control whether or not the subroutine is to remain loaded in core upon completion of the subroutine's execution. If it is omitted, the subroutine will <u>not</u> remain loaded.

PARAMETER=<arg 1>,...,<arg n> is used to pass additional parameters (arguments) to the subroutine. See below.

This command generates the following FORTRAN type call:

```
CALL <subroutine name> (NF,NFIELD,NC,NCAT,NOD,
NDMAT,, <arg 1>,...,<arg n>).
```

Where

NF	is the number of fields in the data set.
NFIELD	is an integer matrix of two dimensions: NFIELD(6,NF). For the Ith field of the data set, NFIELD(1,I) through NFIELD(4,I) is the 16-character field name. NFIELD(5,I) is the number of categories in that field. NFIELD(6,I) is the index into NCAT where those category descriptions begin.

NC	is the total number of categories, i.e., the sum of NFIELD(5,I).
NCAT	is an integer matrix of two dimensions: NCAT(4,NC). For the Ith category, NCAT(1,I) through NCAT(3,I) is the category name. NCAT(4,I) is the category's value.
NOD	is the number of records of data.
NDMAT	is a two dimension integer matrix of data: NDMAT(NF,NOD).
<arg I>	may be an integer, a double word floating point number or a string of up to 24 alphanumeric characters.

(2)

```
CALL <subroutine name> WITHOUT DATA SET
[LIBRARY=<library name>] [KEEP=YES|NO]
[PARAMETER=<arg 1>,...,<arg n>].
```

This form of the call differs only from the above in that all arguments relating to the data set are omitted. This results in the following FORTRAN type call:

```
CALL <subroutine name> (<arg 1>,<arg2>,...,
<arg n>)
```

COMMENTS:

(1) This command does <u>not</u> produce a new RESULT set.

(2) The order in which the keywords appear is arbitrary.

(3) If <arg I> is to be represented as a floating point number, it must contain a decimal point.

(4) The following is a list of acceptable abbreviations for keywords used in this command:

```
PARAMETER | PAR
LIBRARY | LIB
WITHOUT DATA SET | WITHOUT DATASET | W/O
```

(5) Fields containing certain character strings whose length is greater than four are currently truncated to four characters when passed to the called subroutine.

CHANGE

COMMAND DESCRIPTION

PURPOSE: To alter the data in the specified field in either all or certain records of the specified data set.

COMMAND SYNONYM: CH, ALTER, A

PROTOTYPES AND DESCRIPTIONS:

(1) CHANGE IN <data set name> [ALL RECORDS] SUCH THAT <field name> <operand> <new value> [... AND <field name> <operand > <new value>].

Where <operand> can be any of the following:

```
IS
ARE
IS EQUAL TO
ARE EQUAL TO
EQUAL
EQUALS
IS =
ARE =
=
```

and <new value> is a category name, integer value or character string.

This form changes in every record those fields specified by substituting the new value for the existing value.

(2) CHANGE IN <data set name> WHERE <phrase> SUCH THAT <field name> <operand> <new value> [... AND <field name> <operand> <new value>].

Where <operand> and <new value> are the same as defined above in (1). <phrase> refers to any phrase acceptable within the FIND command:

<field name> <verb> <category name>|<value>
[...AND|OR|ALSO WHERE <field name>
<verb>|<category name> <value>].

This form allows the user to find those records meeting the specified criteria and then to change the specified fields (as above in (1)).

(3) CHANGE IN <data set name> WHERE <phrase> TO <new value> [...AND <field name> <operand> <new value>].

This form is identical to (2) above except that the first field name whose value to be changed is not stated following the "TO". It is assumed to be the last field name stated in <phrase>.

COMMENTS:

(1) The RESULT set contains the same number of records as the original <data set name>. However, only those records meeting the specified search criteria will have been changed (i.e., the RESULT set may contain unchanged records).

(2) This command does produce a new RESULT set.

(3) The number of records changed is printed. The percentage of changed records out of all records searched in <data set name> is also printed.

(4) SUCH THAT and TO are synonyms and may be interchanged wherever either is used in the above prototypes.

COMBINE

COMMAND DESCRIPTION

PURPOSE: To combine two data sets into one result set, named RESULT.

COMMAND SYNONYMS: COMB, C

PROTOTYPE AND DESCRIPTION:

COMBINE <data set name 1> WITH <data set name 2>.

COMMENTS:

(1) This command can only be used when both data sets have identical fields. The RESULT is the union of the two sets. (Thus, duplicate records are removed from the result.)

(2) Caution: Currently, the count field of an XTAB set is treated like an ordinary field, thus combining two XTAB sets will result in the removal of duplicate records from the RESULT set. This may mean the loss of information.

(3) This command does produce a new RESULT set.

COMMENT

COMMAND DESCRIPTION

PURPOSE: To add a comment to the output of a MICRO session.

COMMAND SYNONYM: *

PROTOTYPE AND DESCRIPTION:

(1) COMMENT <phrase>.

Where <phrase> may be any character string.

COMMENTS: (1) This command does not produce a new RESULT set.

CROSSTABULATE

COMMAND DESCRIPTION

PURPOSE:		To perform a sorted (ascending, left to right) n-dimensional cross tabulation for the specified fields of a given data set.
COMMAND SYNONYMS:		CROSSTAB, XTAB, X

PROTOTYPE AND DESCRIPTION:

CROSSTABULATE IN <data set name> <field name> [BY <field name> ...] BY [AVE|TOT] <field name>.

Where ... may specify additional BY <field name> phrases.

AVE (average) and TOT (total count) can only appear immediately before the final <field name>.

COMMENTS:

(1) This command does change the RESULT set.

(2) If AVE or TOT is used, then the data referred to by the last field will be treated numerically (instead of categorically).

(3) BY may be replaced by AND or , (comma).

DESCRIBE

COMMAND DESCRIPTION

PURPOSE: To get a description of the specified data set, categories, fields or command names.

COMMAND SYNONYM: D, DES

PROTOTYPES:

DESCRIBE DATA SET <data set name>.

DESCRIBE IN <data set name> FIELD <field name>.

DESCRIBE IN <data set name> CATEGORY <category name> OF <field name>.

DESCRIBE COMMAND <command name>.

COMMENTS:

(1) For a complete listing of all MICRO commands with their descriptions:

```
$COPY SBAU:COMLIST@CC.
```

or secure a copy of the *Technical Reference Manual* from the Institute.

(2) For a listing of just the basic MICRO commands with their descriptions:

```
$COPY SBAU:BASICLIST@CC.
```

(3) This command does *not* produce a new RESULT set.

DESTROY

COMMAND DESCRIPTION

PURPOSE:		To *permanently* remove a data set from disk storage and from access by MICRO.
PROTOTYPE:		DESTROY <data set name>.
COMMENTS	(1)	Once a data set is DESTROYed, it can not be referenced again.
	(2)	After the DESTROY command is issued for a permanent data set, the user will be asked to confirm his action. To reconfirm, the user should type OK. Any other response will result in the command being cancelled.
	(3)	Certain data sets can not be destroyed. If this is attempted, the user will be so informed and no action will take place.
	(4)	This command does *not* produce a new RESULT set.
	(5)	If the data set is to be DESTROYed is a temporary data set, then the command is equivalent to a RELEASE command.

END

COMMAND DESCRIPTION

See STOP command description.

FIND

COMMAND DESCRIPTION

PURPOSE: To extract from a data set those records which match certain specified criteria and to store those records in the RESULT set.

COMMAND SYNONYM: F

PROTOTYPES AND DESCRIPTION:

FIND IN <data set name> WHERE <phrase>.

Where <phrase> may be one or more clauses separated by conjunctions:

<clause$_1$> [... <conjunction> <clause$_n$>].

Where <clause > may be either of the following:

(a) <field name> <verb> <category name> | <value>

Where <verb> is any of those listed in SUPPLEMENTARY INFORMATION on the following page and where <conjunction> is either AND, OR or ALSO WHERE whose meanings are also discussed in SUPPLEMENTARY INFORMATION.

Where <value> may be an integer or character string. If it is a character string, it cannot be greater than 24 characters in length. Any string may be placed in primes ('), but if the string contains embedded blanks (such as 'JONES MOVING COMPANY') then primes _must_ be used.

(b) <field name> IS BETWEEN <value 1> AND <value 2>.

Those elements equal to either <value 1> or <value 2> will also be included.

COMMENTS: (1) The number of records meeting the specified criteria is printed. The percentage of selected records (those in the RESULT set) out of all records searched in <data set name> is also printed.

(2) This command <u>does</u> produce a new RESULT set.

(3) When using a consecutive series of clauses which contain the same field name, the redundant field name need only be stated once. For example:

```
FIND IN JOBS WHERE ZIP IS 48104 OR ZIP IS
48105 OR ZIP IS 48108.
```

Is equivalent to:

```
FIND IN JOBS WHERE ZIP IS 48104 OR 48105 OR
48108.
```

SUPPLEMENTARY INFORMATION:

(1) FIND VERBS (IN EQUIVALENT GROUPS):

```
IS|ARE
IS|ARE EQUAL TO
EQUAL
EQUALS
IS|ARE =
=

IS|ARE NOT
IS|ARE NOT EQUAL TO
IS|ARE NOT =

IS|ARE GREATER THAN
IS|ARE >
>
IS|ARE NOT LESS THAN OR EQUAL TO
IS|ARE NOT EQUAL TO OR LESS THAN
IS|ARE NOT <=
IS|ARE NOT =<

IS|ARE LESS THAN
IS|ARE <
<
IS|ARE NOT GREATER THAN OR EQUAL TO
IS|ARE NOT EQUAL TO OR GREATER THAN
IS|ARE NOT >=
IS|ARE NOT =>

IS|ARE GREATER THAN OR EQUAL TO
IS|ARE EQUAL TO OR GREATER THAN
IS|ARE >=
IS|ARE =>
>=
=>
IS|ARE NOT LESS THAN
```

IS|ARE NOT <

IS|ARE LESS THAN OR EQUAL TO
IS|ARE EQUAL TO OR LESS THAN
IS|ARE <=
IS|ARE =<
<=
=<
IS|ARE NOT GREATER THAN
IS|ARE NOT >

(2) AND, OR and ALSO WHERE

Each <clause> of a FIND command is calculated separately in the order listed in the command statement. The operation defined by each <clause$_i$> is performed on some input set (IN_i) and a result set is generated. In order to avoid confusion between the result set generated by a FIND <clause> and a MICRO RESULT Set, the result generated by <clause$_i$> will be referred to as intermediate set (I_i). The input set for the first clause, <clause$_1$> is the originally named data set, specified in the command by <data set name>. Furthermore after each <clause$_i$> is processed a temporary set is created (T_i) which represents the results of the preceeding i <clause>'s.

There are three basic conjunctions used in the FIND command:

OR

AND

ALSO WHERE

The OR conjunction is somewhat analogous to the union operation in set theory. Specifically, if <clause$_i$> is preceeded by an OR conjunction then the input set IN_i is the same as the input set for the preceeding <clause$_{i-1}$>, IN_{i-1}. Following the operation defined in <clause$_i$>, T_i is formed as the result of a union performed on T_{i-1} and I_i. ($T_i = T_{i-1}$ union I_i). (Note: IN_1 = <data set name>, the originally named set.)

The AND conjunction is somewhat analogous to the intersection operation in set theory. Specifically, if <clause$_i$> is preceeded by an AND conjunction then the input set IN is the temporary set from the previous i-1 clauses,

T_{i-1}. Furthermore $T_i = I_i$.

For example:

FIND IN JOBS WHERE ZIP IS 48104 AND DOT IS 802131 OR DOT IS 802198.

<clause$_1$> is "ZIP IS 48104". Hence, IN_1=JOBS and $T_1 = I_1$.

<clause$_2$> is "DOT IS 802131". Since this clause is preceeded by AND, $IN_2 = T_1$ and $T_2 = I_2$.

<clause$_3$> is "DOT IS 802198." Since this clause is preceeded by OR, $IN_i = IN_{i-1}$ which was equal to T_1 and $T_3 = T_2$ union I_3.

Since <clause$_3$> is the last clause, the MICRO RESULT set would be T_3 which would consist of those records in JOBS where ZIP was equal to 48104 and DOT was either 802131 or 802198.

The ALSO WHERE conjunction is more complicated, but it is somewhat analogous to having a series of FIND and COMBINE commands all in one command. If <clause$_i$> is preceeded by an ALSO WHERE, then <clause$_i$> behaves like <clause$_1$> in that IN_i is the originally named data set. Also, T_{i-1} is saved in a special set S, and T_{i-1} is replaced with an empty set (the null set). When subsequent ALSO WHERE conjunctions are encountered then the temporary set say T_{k-1} is unioned with S and that result is placed in S. Upon the conclusion of the last clause, T_{LAST} is unioned with S to form the RESULT Set.

For example:

FIND IN JOBS WHERE ZIP IS 48103 AND DOT IS 802198 ALSO WHERE ZIP IS 48104 AND DOT IS 802131.

<clause$_1$> is "ZIP IS 48103". Hence IN_1=JOBS and $T_1 = I_1$.

<clause$_2$> is "DOT IS 802198". Since this clause is preceeded by AND, $IN_2 = T_1$ and $T_2 = I_1$.

Since the next clause is preceeded by ALSO WHERE, the previous temporary (T_2) is saved in a special set (S). T_2 is then replaced by an empty set.

$<clause_3>$ is "ZIP IS 48104". Hence IN_3=JOBS (the clause following an ALSO WHERE behaves like $<clause_1>$ in that IN_1 is the originally named data set) and $T_3=I_3$.

$<clause_4>$ is "DOT IS 802131". Since this clause is preceeded by AND, $IN_4=T_3$ and $T_4=I_4$.

Upon conclusion of the last clause $<clause_4>$, T_4 is unioned with S to form the RESULT set.

GET

COMMAND DESCRIPTION

PURPOSE: To acquire the directory information for a group of data sets.

PROTOTYPE AND DESCRIPTION:

GET [DIRECTORIES FOR] <index>.

Where <index> is a character string of up to 16 characters. The index serves as a pointer to the directory information.

COMMENTS:

(1) It is not necessary to GET the directory information for data sets referenced in the user's directory.

(2) The various indices to the different data set groups are available to authorized users through the LMIS Project, Institute of Labor and Industrial Relations.

(3) This command does not produce a new RESULT set.

LIST

COMMAND DESCRIPTION

See PRINT command description.

MTS

COMMAND DESCRIPTION

See SYSTEM command description.

NAME

COMMAND DESCRIPTION

PURPOSE: To temporarily give a different name to any data set.

COMMAND SYNONYMS: N, RENAME, REN, RE

PROTOTYPE: NAME <old data set name> <new data set name>.

COMMENTS:

(1) The <old data set name> will no longer exist after it has been renamed.

(2) The "new name" may be up to 16 characters in length.

(3) Only a temporary data set may be renamed RESULT.

(4) A data set may not be renamed the name of any other data set.

PRINT

COMMAND DESCRIPTION

PURPOSE: To print on a terminal, file or device information about or from a data set, field, category, MICRO commands, etc.

COMMAND SYNONYMS: P, LIST, L

PROTOTYPES AND DESCRIPTIONS:

(1) PRINT [ALL] [DATA] SETS [NAMES].

The names and status of the data sets available to this user are printed.

There are five possible states for the status of a MICRO data set:

(a) DISK - permanent data set on disk; not in core.

(b) DISK* - permanent data set on disk; loaded in core.

(c) TAPE - permanent data set on tape; not in core.

(d) TAPE* - permanent data set on tape; loaded in core.

(e) TEMP* - temporary data loaded in core.

(2) PRINT IN <data set name> [ALL] FIELDS [NAMES].

The names of all fields in the indicated data set are printed.

(3) PRINT IN <data set name> [ALL] CATEGORIES OF <field name>.

The names of all categories in the indicated field in the indicated data set are printed.

(4) PRINT IN <data set name> COUNT.

The number of records in the data set is printed.

(5) PRINT [ON <file name>] [ENTIRE] <data set

name>.

The data for the entire data set is printed (on a file, if specified) record by record.

(6) PRINT IT.

This is synonomous with PRINT ENTIRE RESULT, a specific case of prototype (5).

(7) PRINT [ON <file name>] IN <data set name> <field name 1> [AND <field name 2> ...].

The data of the specified fields is printed (on a file, if specified) for the data set indicated.

(8) PRINT [ALL|BASIC] COMMANDS [NAMES].

The basic list or complete list of command names, and their synonyms, are printed. If neither ALL nor BASIC is specified, then ALL is assumed.

(9) PRINT COST.

The estimated cost of activity since the last COST interrogation is printed. If there was no previous interrogation, then the cost since entering MICRO is printed.

(10) PRINT VM|VMSIZE.

The current virtual memory (core storage) size used by this user is printed.

(11) PRINT TIME.

The current time and date are printed.

(12) PRINT STATUS.

TIME, COST and VMSIZE are printed.

COMMENTS: (1) This command does not produce a new RESULT set.

(2) See Section 1.2.4 for acceptable synonyms.

READ

COMMAND DESCRIPTION

PURPOSE: To read into MICRO a data set which does not have a dictionary and which is in STDS format.

PROTOTYPE AND DESCRIPTION:

READ FROM <file name>.

The dictionary and the data set name of the last explicitly named data set is used with the data read from <file name>.

COMMENTS: (1) This command does produce a new RESULT set.

(2) Always employ the USE command prior to using the READ command. (See USE)

RELEASE

COMMAND DESCRIPTION

PURPOSE: To release the core storage associated with a data set in core.

COMMAND SYNONYM: R, REL

PROTOTYPES AND DESCRIPTIONS:

(1) RELEASE <data set name> [AND <data set name> ...].

The specified data set is released.

(2) RELEASE <data set name>.

(3) RELEASE ALL DATA SETS.

All data sets (both temporary and permanent) which are currently loaded are released.

(4) RELEASE *.

This is synonymous with (3).

(5) PURGE.

This is also synonymous with (3).

COMMENTS:

(1) If the data set is not in core (i.e., not loaded), the command has no effect.

(2) Permanent data sets which are not expected to be referenced again should be RELEASEd in order to reduce the costs associated with the core storage of data sets. This does not, however, preclude further reference to this permanent data set at a later time.

(3) If the data set to be released is a temporary data set, then the RELEASE command has the same effect as the DESTROY command and the data set can not be referenced again.

(4) This command does not produce a new RESULT set.

REMOVE

COMMAND DESCRIPTION

PURPOSE: To extract from a data set those records which match certain specified criteria and to leave in the RESULT set only those records not meeting the specified criteria.

COMMAND SYNONYM: REM

PROTOTYPES AND DESCRIPTIONS:

(1) REMOVE FROM <data set name> WHERE <phrase>.

Where <phrase> refers to any phrase acceptable within the FIND command.

(2) REMOVE <data set name 1> FROM <data set name 2>.

COMMENTS:

(1) This command does produce a new RESULT set.

(2) This command is equivalent to the relative complement concept in set theory.

(3) IN is a synonym for FROM; they may be used interchangeably.

(4) The number of records extracted is printed. The percentage of records extracted out of the total records in the specified set is printed. The number of elements in the result set is also printed.

RENAME

COMMAND DESCRIPTION

See NAME command description.

REPLACE

COMMAND DESCRIPTION

PURPOSE: To replace one data set with another data set.

COMMAND SYNONYM: REP

PROTOTYPE: REPLACE <old data set name> WITH <new data set name>.

COMMENTS:

(1) This command is most likely to be used after a data set has been altered by the CHANGE command.

(2) This command does not produce a new RESULT set.

(3) <old data set name> must refer to a permanent data which is stored on disk.

RESTRICT

COMMAND DESCRIPTION

PURPOSE: To extract those records from one data set whose values of a specified field match those values of a second field in a second data set. These extracted records are placed in the RESULT set.

COMMAND SYNONYMS: RES

PROTOTYPE: RESTRICT IN <data set name 1> WHERE <field name 1> IS <field name 2> IN <data set name 2>.

COMMENTS:

(1) This command does produce a new RESULT set.

(2) It is the extracted records of <data set name 1> that are placed in the RESULT set. If <data set name 1> and <data set name 2> are reversed, then a different RESULT set would be created.

RESTRICTANDMERGE

COMMAND DESCRIPTION

PURPOSE: To merge certain records from two data sets into a single expanded record in the RESULT set.

COMMAND SYNONYM: RAM

PROTOTYPE AND DESCRIPTION:

(1) RESTRICTANDMERGE <data set name 1> BY <field name A> [BY <field name M> ...] WITH <data set name 2> BY <field name B> [BY <field name N> ...].

This form equates specified field names of the first data set with the same number of specified field names of the second data set. Only if the values of the specified field names in the second data set equal those of the first are the entire two records merged (i.e., combined into an expanded record) and placed in the RESULT set.

(2) RESTRICTANDMERGE <data set name 1> BY <field name A> [BY <field name M ...] WITH <data set 2>.

In this form of the command the user only specifies field names for the first data set and MICRO assumes that the second data set has identical field names which are to be used for the comparison. However, the action resulting from this command form is the same as the first form.

(3) RESTRICTANDMERGE <data set name 1> WITH <data set name 2> BY <field name B> [BY <field name N> ...].

This form of the command is similar to the second but the field names are specified for the second data set.

COMMENTS:

(1) This command *does* produce a new RESULT set.

(2) A comparison of the match fields is made of *every* record of the second data set with *each* record of the first data set. If any of the fields to be compared contain unique values

then the number of records in the RESULT set will be less than or equal to the number of records in the larger of the two sets. The number of records in the RESULT set depends of the number of matches and duplicates in the two sets. If there are no unique values in the fields compared, then the number of records in the RESULT set can be greater than the number of records in the larger of the two sets. This can result in extremely large sets due to the combinatorial effect of this situation.

(3) If both sets are cross tabulated sets then the RESULT set will not be a cross tabulated set. If only one of the data sets is cross tabulated then the RESULT set will be a cross tabulated set.

(4) Currently, after a RAM command the RESULT set will not contain the English descriptions for fields and categories of the second data set that could have been printed by the DESCRIBE command. However, all category and field names remain.

SAVE

COMMAND DESCRIPTION

PURPOSE: To permanently save a data set on disk storage for access through MICRO at a later date.

COMMAND SYNONYM: SA

PROTOTYPES: SAVE <data set name> [AS <new name>] [ON <file name>].

COMMENTS:

(1) If a <new name> is not specified, the SAVEd data set is identified by <data set name>.

(2) If the <file name> is not specified, then a new file is created and the new file name will be printed.

(3) If <file name> is specified but no file of that name exists, MICRO will create a file with that name.

(4) This command does not produce a new RESULT set.

(5) If a <new name> is specified, then the data set will be RENAMEd automatically.

(6) MICRO will not allow the user to SAVE a data set with the same name as an already exisitng set.

(7) <file name> may contain up to 16 characters.

(8) The dictionary information for a data set is stored on a file separate from the actual data itself. The procedure for selecting the file name for the data is described above. The name used for the dictionary file is formed by appending a "#" character on the end of the file name for the data, unless that name has 16 characters in which case a "#" character replaces the last character of the data file name.

SELECT

COMMAND DESCRIPTION

PURPOSE: To extract certain fields from each record of a data set and to store the extracted fields as records in the RESULT set.

COMMAND SYNONYM: S

PROTOTYPE: SELECT IN <data set name> <field name 1>
[... AND <field name n>].

COMMENTS:

(1) This command does produce a new RESULT set.

(2) The SELECT command can be used to extract specified fields from a data set, thereby enabling the user to work with a reduced data set. Further, the original data set may be removed from core using a RELEASE command to avoid unnecessary costs associated with core storage.

(3) Caution: SELECT should not be used on a cross tabulated set as the result could be meaningless. Use the XTAB command instead.

(4) Caution: The values of the fields SELECTed should be such that the result is a subset that has no duplicate (identical) records, otherwise duplicates will be deleted without any record of count. This is most easily achieved if the first field name SELECTed contains a unique value for each record in the set. The RESULT set is sorted in ascending order according to the respective order of the fields specified (left-to-right) and duplicate records are eliminated.

(5) Caution: If the purpose of using a SELECT command is to prepare a RESULT set for use with the WRITE FOR ANALYSIS command, care should be taken to ensure that duplicate records are not eliminated because they lack a unique value (key). As stated in (4) above, the duplicates will be deleted and no count will be made. If a count is desired, use the XTAB command to select the desired fields. If duplicates are desired for analysis, the WRITE FOR ANALYSIS command should be used. (See WRITE command.)

(6) AND may be replaced by , (comma) or BY.

SET

COMMAND DESCRIPTION

PURPOSE: To change the action of one of various MICRO Retrieval System facilities.

COMMAND SYNONYMS: TURN, T

PROTOTYPE AND DESCRIPTION:

(1) SET <facility> <status>.

Where <status> is ON or OFF and <facility> is one of the following:

(a) CLOCK (default: OFF)

When the status is ON, MICRO will print after every command the number of seconds of elapsed time and CPU time since the last command. It is initially OFF and when it is first set is ON, it prints the time of day.

(b) ERROR CORRECTION (default: OFF)

When the status is ON, MICRO will attempt to interpret misspelled key words in MICRO commands.

(c) MACRO ECHO (default: OFF)

When the status is ON, any MICRO statement generated by a macro statement will be printed.

(d) ECHO (default: ON)

When the status is OFF, the printing of any remarks following the execution of a command (such as " XX RECORDS IN RESULT SET," etc.) will be suppressed. Error messages will not be suppressed, however.

COMMENTS: (1) This command does not produce a new RESULT set.

SIGNOFF

COMMAND DESCRIPTION

PURPOSE:	To permanently terminate the current MICRO session and to sign-off the computer system (MTS).
COMMAND SYNONYM:	SIG
PROTOTYPE:	SIGNOFF [S|$].
	Where S is the short form and $ is the summary form (only dollar amount used and dollar amount remaining are printed when $ is used).

SORT

COMMAND DESCRIPTION

PURPOSE: To perform an n-dimenseional sort for the specified fields of a given data set.

COMMAND SYNONYMS: SO

PROTOTYPE AND DESCRIPTION:

SORT IN <data set name> <field name> [... BY <field name>].

Where ... may specify additional BY <field name> phrases.

COMMENTS:

(1) This command does produce a new RESULT set.

(2) The number of records sorted is printed.

(3) The RESULT of the command is similar to the RESULT of the XTAB command except that duplicate occurrences are not eliminated and thus there is no COUNT field.

(4) The sort is applied only to those field(s) specified and the RESULT set is ordered accordingly. The unspecified fields are retained, but not used as part of the sort key.

(5) Sort reorders the data set and sorts on the specified fields followed by the unspecified fields in the order that they originally appeared in <data set name>.

STOP

COMMAND DESCRIPTION

PURPOSE: To permanently terminate the current MICRO session.

COMMAND SYNONYMS: ST, END

PROTOTYPE: STOP.

COMMENTS: (1) MICRO cannot be re-entered via a $RESTART command.

SYSTEM

COMMAND DESCRIPTION

PURPOSE: To temporarily leave the MICRO Information Retrieval System and return to the command mode of MTS.

COMMAND SYNONYMS: SYS, MTS

PROTOTYPE: SYSTEM.

COMMENTS: (1) MICRO can be re-entered via a $RESTART command.

(2) This command does not produce a new RESULT set.

TAB

COMMAND DESCRIPTION

PURPOSE: To perform a sorted one-dimensional frequency distribution for the specified field of the given data set.

COMMAND SYNONYMS: FREQUENCY, FREQ

PROTOTYPE AND DESCRIPTION:

```
TAB IN <data set name> [AVE|TOT] <field name>.
```

COMMENT:

(1) This command <u>does</u> produce a new RESULT set.

(2) If AVE or TOT is used, the data referred to by the field name will be treated numerically (instead of categorically).

USE

COMMAND DESCRIPTION

PURPOSE: To make a data set the "last explicitly named data set" as required by the READ command.

COMMAND SYNONYM: U

PROTOTYPE:

```
USE <data set name>.
```

COMMENTS:

(1) This command does <u>not</u> produce a new RESULT set.

WRITE

COMMAND DESCRIPTION

PURPOSE: To write a MICRO data set for future use outside of the MICRO Information Retrieval System.

COMMAND SYNONYM: W

PROTOTYPES AND DESCRIPTION:

(1) WRITE <data set name> [ON <file name>].

This form of the command writes the specified data set in STDS form on the specified file.

(2) WRITE FOR ANALYSIS <data set name> [ON <file name>] [USING <field name 1> [AND <field name 2> ...]].

Where [USING <field name 1> [AND <field name 2> ...]] is similar to the SELECT command in selecting certain fields to be written for analysis. Unlike the SELECT command, however, duplicates are not eliminated.

This form of the command writes the specified data set on the specified file for use with MIDAS at the University of Michigan and CONSTAT at Wayne State University. (MIDAS and CONSTAT are general-purpose statistical programs. See instructions for using MIDAS with MICRO in Appendix A.)

COMMENTS:

(1) If <file name> is not specified, then a new file is created and the new file name will be printed.

(2) If <file name> is specified, but no file of that name exists, MICRO will create a file with that name.

(3) This command does not produce a new RESULT set.

XTAB

COMMAND DESCRIPTION

See CROSSTABULATE command description.

5.2 MACRO SUBSYSTEM

5.2.1 Introduction

The macro subsystem is an extension of the MICRO Command Language. It provides a convenient way to generate a desired sequence of MICRO commands in one or more MICRO sessions. The macro-definition is written only once, and a single command, the macro-command command, is issued each time the user wants to generate the desired sequence of MICRO commands.

An additional facility, called conditional macro generation, allows the user to alter the sequence of commands to be generated during an interactive MICRO session.

It is suggested that this section be read through once and then that Appendix B be referred to for an example of a macro-description and its use.

5.2.2 Macro Libraries

The same macro-definition may be made available to more than one user by placing the macro-definition in a macro library. Once a macro-definition has been placed in a macro library, it may be used by typing its corresponding macro-command during a MICRO session. The procedures used for placing macro-definitions into a macro library will be described in a future update to this manual.

There are two different types of macro libraries. One is the system macro library, the other, a user macro library. Both have the same structure and function. All users have access automatically to the one system macro library during a MICRO session. In addition, when running MICRO the user can indicate which of possibly several user libraries is to be referenced. The user macro libraries contain private macros which may be available only to specified users.

5.2.3 Macro-Definitions

A macro-definition is a set of statements that provides MICRO with:

(1) The name and format of the macro-command, and

(2) The sequence of commands the subsystem generates when the macro-command appears.

Every macro-definition consists of:

(1) A macro-definition name/prototype statement (The DEFINE Statement),

(2) Zero or more delimiter list statements,

(3) Zero or more model statements or conditional macro-generation statements,

(4) A macro-definition end statement. (The % END DEFINITION statement).

A macro-definition cannot appear within a macro-definition, nor can a macro-command. A macro-definition must be available to the system before it's corresponding macro-command is used.

5.2.4 Macro-Command

Macro-commands are commands issued within the main MICRO system. When MICRO recognizes an input line as a macro-command, the macro subsystem is entered. The subsystem, under control of the appropriate macro-definition, generates MICRO commands. The generated commands are then processed like any other MICRO command.

5.2.5 Variable Symbols

A variable symbol is a symbol that is assigned different values by either the user or the macro subsystem. When the macro subsystem interprets a macro-definition, variable symbols in the model statements are replaced by values assigned to them. By changing the values assigned to a variable symbol, the user can vary the contents of the generated commands.

There are three types of variable symbols: symbolic parameters, symbolic delimiters, and system variable symbols.

Symbolic parameters are written with an at-sign (@) prefix, followed by one to two digits in the range 1-99. Symbolic parameters are assigned values by the user each time he writes a macro-command. The subsystem will accept as valid any such value; however, the value put in the generated command may be rejected as invalid by the main MICRO system.

Symbolic delimiters are written with an at-sign prefix followed by the letter D followed by one to two digits in the range 1-99. Like symbolic parameters, symbolic delimiters are assigned values by the user each time he writes a macro-command. However, the subsystem will accept as valid values only those that are listed in the macro-definition.

System variable symbols are assigned values by the subsystem each time it processes a macro-command. Currently, there is one system variable. It is written as @NULL and is assigned the value of a null string.

5.2.6 Writing Macro-Descriptions/Commands

A macro-command can have any number of lines. It is terminated by a period (.). A macro name/prototype statement can also have any number of lines and is also terminated by a period. Model statements are only one line long. Conditional generation and delimiter list statements are also only one line long and the first character of the line must be a percent sign (%). The macro-definition end statement consists of only one line and the first character must be a percent sign.

In this section, the same notation is used as in the main section of this publication.

DEFINE

MACRO SUBSYSTEM STATEMENT

PURPOSE: To indicate the beginning of a macro-definition and to specify the macro name and the format of all macro-commands that refer to that macro-defintion.

SYNONYM: DEF

PROTOTYPE: DEFINE <macro name> [<delim 1>] [<param 1>] [[<delim m>] [<param n>] ...].

Where <delim m> is a symbolic delimiter of the form: @Dn

and <param m> is a symbolic parameter of the form: @m

COMMENTS:

(1) This statement must be the first of every macro-definition.

(2) The macro is given the name <macro name>.

(3) The presence and position of symbolic delimiters and/or symbolic parameters indicate where the actual delimiters and parameters appear in the macro-command.

(4) The prefix character is changed to =.

EXAMPLES:

```
DEFINE  PUT.

DEFINE LOOK AT @D1 @1 @2.
```

DELIMETER LIST

MACRO SUBSYSTEM STATEMENT

PURPOSE: To specify the list of values that a symbolic delimiter may be assigned by a macro-command.

PROTOTYPE AND DESCRIPTION:

```
%D<n> (<char string 1> [,<char string m> ...])
```

The nth symbolic delimiter is assigned the list of values in parentheses.

COMMENTS:

(1) Delimiter list statements must immediately follow a DEFINE statement and precede any model or conditional generation statements.

(2) Not every symbolic delimiter used in a description need be given a list since every symbolic delimiter is assumed to have a comma (,) as an allowable value.

EXAMPLES:

```
%D1  (BY,AND,WITH)
```

END DEFINITION

MACRO SUBSYSTEM STATEMENT

PURPOSE: To indicate the end of a macro definition.

PROTOTYPE:

```
%[<label>] END [DEFINITION]
```

COMMENTS:

(1) This statement must be the last of a macro-definition.

(2) The macro-subsystem is exited and the MICRO system is re-entered (and the prefix character reverts to a -).

EXAMPLES:

```
%  END DEFINITION

%OUT END
```

GOTO

MACRO SUBSYSTEM STATEMENT

PURPOSE: To unconditionally branch to another macro statement within the current macro definition.

PROTOTYPE: %[<label 1>] GOTO <label 2>

Macro processing will continue at the statement labeled <label 2>.

COMMENTS: (1) This is a conditional generation statement.

EXAMPLES:

```
% GOTO OTHER

%HERE GOTO TEST2
```

IF

MACRO SUBSYSTEM STATEMENT

PURPOSE: To test the value of a symbolic variable and, depending on that value, process either the next macro statement or one elsewhere in the macro-definition.

PROTOTYPE:

```
%[<label 1>] IF <sym var 1> EQ|NE <char
string>|@NULL|<sym var 2> GOTO <label 2>
```

where <sym var> is any symbolic delimiter of the form @Dn or symbolic parameter of the form @n.

If EQ is specified, then if the value of <sym var 1> equals @NULL or <char string> or <sym var 2> then the macro statement labled by <label 2> is processed next. Otherwise the next macro statement will be processed.

If NE is specified, then the condition tested for is inequality.

COMMENTS: (1) This is a conditional generation statement.

EXAMPLES:

```
% IF @1 EQ PLOT GOTO THERE

%BYTEST IF @D2 NE @NULL GOTO OUT
```

MACRO-COMMAND

MACRO SUBSYSTEM STATEMENT

PURPOSE: To start the generation of MICRO commands according to a specified macro-definition.

PROTOTYPE: <macro name> [<value delim 1>] [<value param 1>] [[<value delim N>] [<value param M>] ...]

COMMENTS: (1) The definition named <macro name> is invoked and starts generating MICRO commands. The values specified by the macro-command replace the corresponding symbolic variables in the definition.

(2) As each MICRO command is generated it is processed.

(3) Normally the MICRO commands generated are not printed. See the MICRO SET command described previously on how to effect printing of the generated commands.

Examples: DISPLAY JOB DESCRIPTIONS.

MODEL

MACRO SUBSYSTEM STATEMENT

PURPOSE: To specify the text to be used in generating a MICRO command.

PROTOTYPE: Any sequence of blanks and/or characters including symbolic delimeter and/or parameter variables.

COMMENTS: (1) The text is put, as is, into the MICRO command being generated. However, any symbolic delimiters or symbolic parameters are replaced by the actual values in the macro-command.

EXAMPLE: FIND IN @1 WHERE @2 @D5 28

NOP

MACRO SUBSYSTEM STATEMENT

PURPOSE:		To provide a point of reference for conditional generation of MICRO commands.
PROTOTYPE:		`%[<label>] NOP`
COMMENTS:	(1)	The subsystem just passes over NOP statements.
EXAMPLE:		`%THERE NOP`

APPENDIX I-1
MIDAS

Michigan Interactive Data Analysis System (MIDAS) is a statistical package available on MTS at the University of Michigan only.[2] The MICRO command

```
WRITE FOR ANALYSIS <data set name> [ON <file name>].
```

creates a file with the same name as <data set name> (or if ON <file name> is specified, it creates a file with that name). This information can be accessed at any time *after* the current MICRO session by issuing the following command

```
$SOURCE <file name>
```

where <file name> refers to the file created by the WRITE FOR MIDAS command. If <file name> is a temporary file, it can only be accessed after leaving MICRO and prior to $SIGning off the system.

Once you have $SOURCEd the file and entered MIDAS, it should be noted that:

(a) All field names in the data set are MIDAS variables;

(b) These variables are MIDAS analytical variables which are real, double precision numbers.

---

[2] See *Documentation for MIDAS*, Statistical Research Laboratory, University of Michigan, 1972, 126 pp.

APPENDIX I-2
MACRO EXAMPLE

The following is a sample macro-definition:

```
DEFINE GIVE @D1 @D2 @1 @D3 @2.
%D1 (STATISTICS, RANGE, MOMENTS)
%D2 (FOR,OF)
%D3 (IN)
SELECT IN @2 @1.
CALL TALYHO KEEP=Y
% IF @D1 EQ STATISTICS GOTO THREE
% IF @D1 EQ RANGE GOTO ONE
    PAR=2.
% GOTO END
%THREE NOP
    PAR=3.
% GOTO END
%ONE NOP
    PAR=1.
%END END DEFINITION
```

The macro GIVE can be invoked by any of the following forms of macro-instructions:

(1) GIVE RANGE OF <field name> IN <data set name>.

(2) GIVE STATISTICS FOR <field name> IN <data set name>.

(3) GIVE MOMENTS OF <field name> IN <data set name>.

(4) GIVE STATISTICS OF <field name> IN <data set name>.

The form (1) would cause the generation of the MICRO commands:

```
SELECT IN <data set name> <field name>.
```

```
     CALL TALYHO KEEP=Y PAR=1.
```

While forms (2) and (3) would produce:

```
     SELECT IN <data set name> <field name>.
     CALL TALYHO KEEP=Y PAR=3.
```

And form (4) would produce:

```
     SELECT IN <data set name> <field name>.
     CALL TALYHO KEEP=Y PAR=2.
```

APPENDIX J

THE LMIS DATA BASE VERSION 2

Michael A. Kahn and Boyd L. Bronson

6.0 Introduction

The Labor Market Information System (LMIS) Project's data base consisted of numerous standard data files dealing with various aspects of labor market information collected by different levels of government. The following data files were included:

Census of Population
Current Population Survey
EEO-1
ESARS
ES202
ES203
Job Bank
Social Security
Urban Employment Survey

Each data file had been adapted for use with MICRO, the LMIS interactive information retrieval system [8].

A description of each data file that was in the LMIS data base follows. Then, in Appendix J-1, a listing of the data files and the details of the location of the data sample is presented. A paper containing a detailed listing of the categories of information for each data file accessable through MICRO is available on request.

6.1 Census of Population

The Bureau of the Census released six one percent samples of the 1970 census. Three of these public use samples contain micro data records from a questionnaire completed by five percent of the population and the other three contain micro data records from a different questionnaire presented to fifteen percent of the population. There are three geographic breakdowns for each questionnaire:

1) State public use samples
2) County group public use samples

3) Geographic division public use samples with neighborhood characteristics.

The LMIS data base is incorporating data from both state and county group samples for both questionnaires. (See Appendix J-1 for a detailed description of the areas covered by Census information in the LMIS data base).

6.2 Current Population Survey

The Current Population Survey (CPS) is a monthly survey conducted by the Census Bureau of approximately 50,000 occupied households. The sample includes 449 samples areas, covering every state and the District of Columbia. Information for more than 100,000 persons fourteen years of age and over is collected every month in the survey. The survey is designed to provide individual and family information, both from the March tapes. The major limitations of CPS are twofold. First, the sample size is limited for our project's SMSA's. There are only 2,000 individuals in the sample of the project's three SMSA's combined. Second, the data are subject to errors due to the failure of the respondent to remember correctly and his intentional misinformation to exaggerate the prestige of an occupation or income.

6.3 EEO-1

Special arrangements were made with the Equal Employment Opportunity Commission (EEOC) to provide the LMIS Project with the EEO-1 responses for the Denver, Detroit, and Milwaukee SMSA's. Tabulations from this data set which meet EEOC confidentiality requirements are shown in Appendix M.

Filing of reports was required by the Civil Rights Act of 1964 for all employers with 100 or more employees and certain other government contractors. Tallies are based on visual counts of employees by the employer. Government employees are exempted from the reporting. In addition, EEOC estimates that only 75% of the employers required to respond actually did respond in 1967. However, 95% of all large employers responded. There is a time delay of one year in the release of data.

6.4 ESARS

Employment Service Automated Reporting System (ESARS) is an attempt to automate the reporting of transactions for the accounting of activities and accomplishments of the public employment service offices. ESARS was developed to enable planners to get the type of information from the Employment Services, extremely detailed information, which can only be handled efficiently by an automated system. ESARS is based on individuals rather than transactions, a basis for more effective management information systems in manpower programs. There are two data files for ESARS in the LMIS data base - Applicant Characteristics and Job Orders.

Confidentiality requirements were strictly adhered to with ESARS as with other data sets. The only information linking individuals with records in the Applicant Characteristics file was the social security number.[1] This was removed before the data was put into its final form for use with the MICRO language. The only exception was used at the Denver Youth Opportunity Center (YOC) where YOC applicant information was made available for applicant searches. The Job Order file contained no information identifying the employer.

The major disadvantage of ESARS data is the quality of the input and the low coverage of the Employment Service. The former should be corrected with the implementation of Manpower Operations Data Systems (MODS).

6.5 ES202

ES202 is a data file derived from a mandatory employer report submitted to each state's Employment Security agency by every establishment in the state covered by unemployment insurance. The report is used primarily for the administration of the states' Unemployment Insurance law, the Bureau of Economic Analysis personal income estimates and drawing samples for the Bureau of Labor Statistics Surveys. As a result of 1970 Employment Security Amendments, approximately 65 million jobs are covered by Unemployment Insurance. Excluded are 12 million jobs.

---

[1] ESARS data can be accessed only by project staff and the respective state Employment Security offices.

Two-thirds of those excluded are in state or local government. The remainder are in domestic service, agriculture, small firms and non-profit organizations [16].

Each state has its own system of collecting the data and can require different information in its report. Common data items collected in the different states include monthly employment of workers covered by Unemployment Insurance, quarterly wages and Unemployment Insurance liability by establishment. Data can be cross-classified by industry or county. Data are available within six months after the end of the quarter being reported.

Confidentiality requirements vary from state to state. Data were "scrambled" in the computer to prevent unauthorized access.[2] Each participating state has reviewed the procedures to be sure that their confidentiality requirements are met.

6.6 ES203

ES203 is a report on the characteristics of the insured unemployed. The combination of ES203 data and Social Security data would result in information on both the unemployed and the employed sectors of the population, within the scope of these two samples.

6.7 Job Banks

The Job Bank program is intended to serve a multitude of purposes related to the processing of job information. The Job Bank System was not implemented in all areas, consequently the data for Job Banks files was only available from certain areas. The LMIS project's three SMSA's were included in the Job Banks program and four data sets for Denver were prepared for use in the computerized aid to counseling project at the Denver Youth Opportunity Center.

Information in Job Bank that was not available in ESARS included minimum pay and rate of pay information and more detailed information about job openings. Confidentiality precautions for Job Bank were the same as for ESARS.

---

[2] Only the state Employment Security representatives and LMIS personnel had access to these files.

6.8 Social Security

The Regional Economics Division of the Office of Business Economics provided three tabulations based on the one percent continuous work history file. The first tabulation included information on the number of individuals cross-tabulated by sex, age, 1970 industry (three digit standard industrial classifications), 1970 sub-SMSA (standard metropolitan statistical area), 1970 wage and 1965 work history status. The second tabulation included information by job holder rather than individual by wage, industry, sex, race, and age. Thus in the second tabulation each person was counted as many times as he had covered jobs. This permits some comparability with ES202 data. The third social security tabulation compared the job history in 1971, 1970, 1965 and 1960 of covered individuals working in Colorado, Wisconsin, Michigan, Wyoming, Montana and Utah during the four selected years.

The major disadvantages of these data were the absence of non-covered groups such as federal employees, absence of occupational detail and the crude method of computing annual wages.

6.9 Urban Employment Survey

During 1968-1969 the Bureau of the Census conducted a survey of characteristics of individuals in the Concentrated Employment Program areas (CEP's)[3] of six cities: Atlanta, Chicago, Detroit, Houston, Los Angeles, and New York. The Census Bureau also surveyed the non-CEP areas of Detroit and Atlanta. A sample of 3,500 households was drawn from each of these eight areas.

Because of confidentiality restrictions, micro data could only be obtained for the non-CEP area of Detroit. The survey consists of data on population characteristics, unemployment, work experience, earnings, family income, educational attainment, occupation and industry, as well as how a worker found a job. Tabulations for the CEP area were published by the Bureau of Labor Statistics in "Poverty - The Broad Outline - Detroit", Urban Employment Survey - Report #1.

---

[3] Concentrated Employment Program Areas refer to target areas in which the U.S. Department of Labor has combined separate manpower programs in order to concentrate the impact of these programs.

APPENDIX J-1

DATA SETS IN THE LMIS DATA BASE

December, 1973

Census (1970) - Person Information Only

- Colorado - 5% Sample
- Denver SMSA - 5% Sample
- Milwaukee SMSA - 5% Sample
- Montana - 5% Sample
- North Dakota - 5% Sample
- Oakland and Macomb Counties - 5% Sample
- South Dakota - 5% Sample
- Utah - 5% Sample
- Wayne County - 5% Sample
- Wyoming - 5% Sample
- Colorado - 15% Sample
- Denver SMSA - 15% Sample
- Milwaukee SMSA - 15% Sample
- Oakland and Macomb Counties - 15% Sample
- Wayne County - 15% Sample

Current Population Survey 1970 - Employment Characteristics

- Wisconsin

Current Population Survey 1970 - Miscellaneous Characteristics

- Wisconsin

EEO-1 1970

- Denver
- Detroit
- Milwaukee

ESARS Applicant Demographic Characteristics

- Denver YOC 1973
- Milwaukee June 1971

DATA SETS IN THE LMIS DATA BASE continued

- ESARS Job Orders
    - Detroit June 1971
    - Milwaukee July 1970 - June 1971
- ES202
    - Denver (1969 - annual)
    - Milwaukee (1970 - 1st quarter)
    - Detroit (1970-1971, 5 quarters)
- ES203
    - Milwaukee - 1971
- Job Bank 1973 - Denver
    - Jobs
    - Job Related Services
    - Non-Job Related Services
    - Referrals
- Social Security Multiple Job Holders 1970
    - Denver
    - Detroit
    - Milwaukee
- Social Security 1967
    - Denver
    - Detroit
    - Milwaukee
- Social Security 1970
    - Denver
    - Detroit
    - Milwaukee

DATA SETS IN THE LMIS DATA BASE continued

Social Security 1971

- Colorado
- Michigan
- Montana
- Utah
- Wisconsin
- Wyoming

Urban Employment Survey 1968

- Detroit

APPENDIX K

MEETING THE NEEDS OF USERS OF A LABOR MARKET INFORMATION SYSTEM

Malcolm S. Cohen and Arthur R. Schwartz

7.0 Introduction

For the past three years the Labor Market Information System Project at the University of Michigan has carried out a feasibility study for the Department of Labor to help the Secretary of Labor develop a comprehensive labor market information system (LMIS), as called for in Title III of the Comprehensive Employment and Training Act. The Secretary of Labor was directed to "develop a comprehensive system of labor market information on a national, state, local or other appropriate basis."

The University of Michigan (U-M) received one of several university contracts awarded by the Manpower Administration. The emphasis in this contract was on an analysis of the various data bases now available, their gaps and limitations and how information from these data bases can be made more readily available. The contract also called for an analysis of conceptual needs of users of these data bases.

7.1 Needs of Users

The purpose of this appendix is to present a short summary of the needs both met and unmet of the major users of the labor market information system (LMIS) and costs associated with these needs. Since this is meant to be a summary, there will be many very interesting issues that will have to be left untouched. But, it is hoped that this paper will generate some agreement on the gaps in the labor market information system as well as point out needs of users that are being met.

Other investigators have studied needs for a LMIS.[1] Our study does not attempt to either summarize these studies or duplicate them. Our interest lies primarily in setting forth a description of needs of users which can or can not be satisfied with existing data bases.

When evaluating the need for a labor market information system, one should begin by asking the questions: 1) who will use it and 2) what specifically will each user need from this system? The users can be specified by an analysis of the labor market: its inputs, outputs, and everyday operations. To best analyze the needs of each particular group of users, it is best to go directly to the source. The analysis of user needs is based not only on the opinions of those people familiar with the broad topic of labor market information, but also on interviews conducted with the actual participants in the daily operations of the labor market.

---

[1] See, for example, Yavitz and Morse [20] and Margaret T. Larsen [15].

A distinction is sometimes made between management information and labor market information. Management information includes information necessary for the management of manpower programs, including the Employment Service. In practice the distinction is difficult to make.

The needs of all major users of labor market information are discussed in this appendix. However, our experimental data base is designed for use of the Research and Analysis Departments of Employment Services and individuals who have responsibility for the quality of the data bases. These users have the most sophisticated knowledge of the data bases and are aware of their limitations. Others, aware of the data limitations, can also benefit from the data bases. Designing the experiment for any possible user is far too costly at this stage. However, since an ultimate objective of labor market information is to help the worker find a job, this concern is reflected in this appendix.

7.2 The Users and Their Needs

There are three types of persons that would use a labor market information system (LMIS). These are workers (including unemployed job seekers, job changers and first-time job seekers), employers, and the group of people such as counselors, planners, and government officials who must make decisions and give advice based at least in part on information that they have received from the LMIS. There are different subgroups of each basic group, each with its own particular needs.

Workers: The primary informational need of workers is for job information. However, the worker is not only a user of labor market information, but he is also a potential provider of inputs into the system. There are two types of data that workers can provide. The first is operational data. An example of this is the application form filled out by all workers who seek help at the State Employment Service. If a worker applies, he generally fills out this form. The worker's perception of the Employment Service's ability to help him is the key to his providing this input. The worker can also provide information through surveys. The input here will depend on the success of the government collection agency.

Because of the complexity of the labor market, it is better to make a finer division of the workforce. The main subgroups would be: 1) job-ready employed, 2) job-ready unemployed, 3) non-job-ready unemployed, and 4) persons entering the labor market for the first time. For all of these groups the primary need is for job search information, but for each one these is slight variation. "Job Search Information may be defined as information which will assist an applicant or UI climant to obtain a suitable job or training opportunity consistent with his aspirations and qualification..." Chavrid, [2], pp. 16-17.

1) The job-ready employed worker would not usually be looking for a job, but presumably he would take a "better" one if he knew it were available. In a study dealing with changes of workers from blue collar to white collar jobs, Jobin and Stern [10] found that the most important souces of information to workers changing jobs were: (1) friends and relatives and (2) newspaper ads and direct application. There was little or no reliance on the Employment Service. This may be due to the workers' belief that the Employment Service has little to offer them. This group would probably make more use of the Employment Service if a very complete listing of job openings were available to them. The listings would also have to have sufficient detail on the nature of the jobs, pay and qualifications.

2) The job-ready unemployed workers make up a group that the LMIS can better serve. Their interest is in job listing, but they also want detailed information about the jobs. They want to know what the wage scales are, what the necessary qualifications for the jobs are, and many other details about each job opening that will better help

them to select a job that will fit their skills and desires. They are also interested in area supply and demand information for their area and occupation. This group of workers will usually come to a central agency such as the Employment Service, if they perceive a benefit. For this reason the Employment Service must convince the work force at large that it has good job listings. Of course, the Employment Service must also convince employers to list good openings. However, in many cases the Employment Service has not been able to do either. While an unemployed person must register with the Employment Service to collect Unemployment Insurance, this does not mean the Employment Service can help him find a job. Some occupations have better listings than others. There may be no job opening at all listed at the Employment Service compatible with the worker's past experience or training. Thus, the worker turns to other sources to find a job. In a study of five cities hit by major plant shutdown, Wilcock and Frank [19] found a small utilization of the Employment Service as Table 1 indicates:

TABLE 1

HOW JOBS ARE FOUND
(Figures are percentages)

	E. St. Louis	Columbus	Fargo	Oklahoma City	Peoria
Friends and Relatives	53	37	31	33	43
Direct Application	22	32	35	40	31
State Employment Service	3	4	9	4	5
Company or Union	7	12	7	3	5
Other (Want ads, Private Agencies, etc.)	15	15	18	20	16

Source: Wilcock and Franke [19], p. 129

It appears that the workers did not perceive that the State Employment Service could be of very much help to them. We hypothesize that this is because of the paucity of job listings at the Employment Service. Only a small fraction of jobs in the community are listed in the Employment Service. An estimated 15-20% of all job vacancies are on file with the Employment Service.[2] Alternatively, the description of the jobs may not be as detailed as a friend could provide. Furthermore, the jobs may not be the most attractive jobs available in the community. All these factors contribute to a bad image for the Employment Service. Even if the job listings improve, the image can linger on.

3) The non-job-ready unemployed worker is the major target of many manpower programs. This group is the hardest to reach. It is hard to know exactly what they want, and if one could find these wants, it would be hard to get the information to them. Often these are ghetto area workers who "have an unrealistic view of wages available to untrained and inexperience entry workers, and are looking for 'instant jobs' or 'career jobs' at high pay without recognizing the need for advanced training and education to prepare themselves for these jobs." (Chavrid [2], p. 19).

This type of worker first needs extensive vocational counseling and training, and then job listings. He must be provided with career information including occupational requirements, information on where to acquire particular skills (i.e., vocational education schools), and then finally, where the jobs can be found. The needs that are primarily unmet are good knowledge of private vocational education schools and good occupational demand data. Basic career information issued in usable form should be made available especially to this group of workers.

4) A worker entering the labor market for the first time is looking first for career and occupational guidance. In a study of guidance sources Edward Kalachek [11] found a surprising dearth of guidance for high school dropouts. (See table 2).

It appears that the problem is again that of being unable to reach those who need it most. The young worker who does go for guidance is looking for occupational supply and demand projections, job qualifications literature, and information about training. Most of the same unmet needs that apply to the non-job ready workers apply to this group of workers. One difference would be that the young workers are looking more for a long range outlook or perhaps a wider area survey of supply and demand since they are presumed to have greater mobility. They are willing to change jobs to find their "calling".

---

[2] Arets, [1], p. 124. In manufacturing alone, employment service placements were 16.3% of all new hires for the period Jan-67 to Nov-69.

TABLE 2

SOURCE OF JOB GUIDANCE
(Figures are percentages)

	Dropouts		Graduates (High School)	
Received guidance	22.4		56.1	
School Only		17.1		37.8
Employment Service Only		4.2		4.9
School and Employment Service		1.0		13.4
No guidance	77.6		43.9	

Source: Kalachek [11], p. 85

Employers. The employer is not only a major user of LMIS, but is also a potential supplier. He is most likely to supply information if he perceives a benefit of doing so or in the case of law, a penalty for not doing so. One very important potential input is job vacancy information. To provide the job information the employer must be convinced that he will get "good quality" workers for his effort. If the employer feels he will have to do expensive screening of applicants sent to him by the Employment Service, to find a single qualified applicant, employers are going to be less likely to send job orders to the Employment Service. As with applicants there may be a lag between employer use of the Employment Service and improvement in Employment Service performance.

Another potential data input is the employer reports to government agencies, such as the mandatory Social Security and Unemployment Insurance Programs, the voluntary Job Opening and Labor Turnover program, and the Occupational Employment and Employment Statistics programs of the Bureau of Labor Statistics.

One type of need that employers often mentioned was that of information for future planning. For example, a company planning to relocate a plant would want information about the areas it had under consideration. It would be interested in industry and occupational wage rates in that area. Also, it would be interested in the present and future labor supply. What is the unemployment rate; what is the skill mix; and what is the area that workers can be drawn from? These are some of the questions the employer wants answered. All of these data would have to be on a local basis. To meet these needs, better demographic and occupational data are necessary. Other needs that were mentioned are current industry and occupational wage data to indicate to a given employer what the conditions are in certain industries. Also good occupational supply and demand data was desired. They mentioned that extensive area supply data were limited to the Decennial Census and that present demand and wage materials had many gaps in them. Another complaint that employers had was that the material was not specific enough. For example, some employers criticized the Job Opening and Labor Turnover (JOLTS) form because its published occupational information is too broad or missing entirely.

It is very important that employers have an accurate picture of labor market conditions in a given area. Employers who read about high unemployment, and yet are unable to fill their own vacancies will be very skeptical about actual labor market conditions. (See Chavrid [2], p.23) So the need here is for more detailed information on or perhaps better measures of vacancies, shortages and surplus labor by occupation, industry and area.

Counselors and Planners. The final users of the LMIS to be discussed are those who make decisions and give advice based on information they receive from the labor market. These are the people who look at the labor market from the outside, analyze what they see, and make decisions which could affect an individual, a group, or an entire governmental agency. These are the counselors both in schools and Employment Services, and the planners and decision makers both in private industry and government. The school counselors deal primarily with those entering the labor market for the first time, while the Employment Service counselors deal with a wide variety of people: those entering the labor market for the first time; those re-entering it in a new field;

those who are unemployed and those who are just confused and have nowhere to go. Planners may be interested in cost-benefit analysis for different programs or they may be interested in projections to evaluate some decision with future implications. Finally the decision makers of government agencies such as heads of Employment Services can make use of management information data to aid them in more effectively overseeing the operation of their various agencies. This might be called the need for internal operating information.[3]

Counselors need information about where the jobs are, what the jobs offer (wages, etc.), what kinds of training are needed and available, as well as information about the future outlook of this particular industry or occupation. The single greatest complaint of school and Employment Service counselors was the lack of integration of data. A great deal of information would come into their office, but it would be in piecemeal form. (Hence, the need for a Labor Market Information System). Too often the information would be too general for practical use. For instance, a counselor may get information about the building trades, but it will usually be descriptive in content and in outlook. It will be very hard to know where the jobs are, what the pay and requirements are, and what the future demand will be. Once again, it is the problem of poor occupational data, which in the case of a worker entering the labor force for the first time is especially vital.

The counselor should also know something about the vocational education programs. Such information might include: what kinds of programs are offered; what programs lead to what occupations; the probability of completing the program and the probability of finding a job after completing the program. Although there is information available for public vocational education programs, there is virtually nothing available for private vocational education, and, therefore, a very definite gap.

The needs of planners will be assessed in two contexts: (1) cost-benefit analysis, and (2) forecasting-planning decisions. The questions most relevant to cost-benefit analysis are: (a) What is the universe of need? (b) What will it cost to serve the universe of need? (c) How have enrollees benefited from the program? (d) What are alternative programs? (e) What enrollees are most likely to benefit? There is often the need for individual cost benefit analyses for different demographic groups as well as good follow-up studies. What was the age breakdown of those who finished the manpower programs? What is the racial breakdown of those who could not be helped by the Employment Service? These are some of the questions that could be answered with the right data. Follow-up studies are done but they are extensive enough. For example, there is little cross-classification between success ratios in vocational education and race.

---

[3] This is discussed in the next secion of this paper -- Management Information.

The people who make projection planning decisions, need "future information". What will be the total supply inflow into the labor force in a certain area in the next five years? What will future demand conditions be? These are very difficult questions to answer. A logical start is to identify the major potential flows into the system: those entering the labor market from regular and vocational schools, the number of workers migrating into the area, the number of workers being trained and promoted minus retirements and deaths and outmigrants. There must be conceptual relationships derived for the flows. If "x" number of students are enrolled in physics in college, how many future physicists will come from this group? What is the relationship between vocational education enrollment and completions? Is there any difference in probable success and placement depending on the type of program that worker attends? Will completion necessarily mean a new input into the labor market and what field will it be in? What is the rate that the present workers will leave their particular occupations? (See Goldstein [7]) for a description of a Bureau of Labor Statistics program to estimate attrition by occupation). These are questions relevant to the projection-planner. Data limitations exist in all of these areas but are most serious for vocational education, especially private vocational education schools. The general school enrollment question is difficult to answer also, but at least there is some relevant data available. The conceptual relationships have not been developed, but some supply data exists.

The information for national decision-makers is needed to help them determine how effective the existing manpower programs are, and how future programs should be directed to be most effective for a given amount of expenditures or, more commonly, how expensive a program will be to serve a particular universe in need of manpower services. On another level, heads of State Employment Services need better information to help them in the operation of their agency. Questions as to the success of operations, who is being helped, and how activities could be better organized to serve those who need it the most need to be answered. Only with better and more frequent demographic data a greater volume of follow-up studies can these questions be answered completely. The new ESARS (Employment Service Automated Reporting System) program will better unify information on Employment Service activities and hopefully generate more useful and accurate data.

Conceptual models connecting labor supply and demand for present needs and future planning are needed both on the national and local levels of planning. For example, if one is to evaluate the need for a manpower training program or a vocational education program, many questions will be asked that cannot be answered with existing data alone. For example: (1) How many people will be trained? (2) What will be the characteristics of the successful trainees? (3) What are the chances of a successful trainee finding a job? (4) What will the addition of new trainees do to the unemployment rate in a given area? (5) How do changes in the unemployment rate affect new hires, labor turnover, or job orders to or placements by the Employment Service? (6) What can be done to correct excess supply or demand in a given industry or occupation in a particular area? These are problems that planners at all levels have to deal with for manpower programs of any kind. Even with good data these kinds of questions require good conceptual models. However, good models and data are not enough. The decision maker using the data or model must combine his knowledge of certain decision variables with the data and model for good decisions to be reached.

Several new statistical programs will help to fill in some of the former gaps in the LMIS. First, the new Occupational Employment Statistics (OES) program should help considerably with the many gaps in occupational information, especially on the supply side. The new Employment Service Job Banks Program will help provide additional job listings and perhaps will also help in meeting some of the occupational demand needs. The Post Census Employment Survey will provide much more detail on the characteristics of residents of the nation's poverty areas as they pertain to employability.

7.3 Management Information

In the previous section the needs of decision makers were discussed. This can be classified as management information. However, in addition, persons managing the employment service have information requirements. The major need as we perceive it is to provide a relationship between the cost of programs and their benefits.

The cost of programs might be measured by data currently collected in an automated accounting system. The benefits might be measured by the number of individuals served and their status in time after being served.

The Employment Service Automated Reporting System (ESARS) provides data on individuals served. However, the cost of collecting the data is very large and thus has put a great burden on state agencies. A number of state agencies have complained both about the burden and the quality of data. A frequent complaint has also been that the reports generated by ESARS do not meet their needs.

A large gap in management information exists because of the failure of any of the reporting systems to provide data on the individuals either missed by the Employment Service, persons who drop-out of the service records and persons who are served by the Employment Service but never show up again.

This large gap can be narrowed in two ways. First, states should have funds for follow-up surveys and household surveys which will permit analyses to be made of persons not appearing in Employment Service records. Second, attempts should be made to link in a comprehensive information system non-employment service labor market information with Employment Service records. Such an attempt is underway at Michigan.

One possible way to cut down on the cost of ESARS reporting would be to substitute a short form for applicants not in computer matching states and only require a full ESARS report for a sample of applicants.

Another possible way to improve the value of ESARS' information is by replacing ESARS reports with an online data system. It is almost impossible for a cross tabulation program to satisfy all the varied needs of the users.

APPENDIX L

AN INDEX TO MAJOR PUBLISHED DATA ELEMENTS FOR USERS OF LABOR MARKET INFORMATION

Arthur R. Schwartz and Malcolm S. Cohen

8.0 Introduction

This appendix is an updated version of a working paper written in 1971. This appendix is not intended to be an all-inclusive directory of sources of labor market information. What it is meant to be is a compact source of the primary reports that make up the labor market information system (LMIS). Knowledge of these sources would enable one to have a good grasp of the foundations of the LMIS. It would have been easy to simply list a tremendous number of publications for each of the departments mentioned. However, this would have produced a very cumbersome document. Instead, the most important reports for each agency were selected, those that seemed most relevant to the general needs of the users of the LMIS.

This report is divided into two parts. The first is a short summary of each report. The second part is comprised of a simplified "information matrix". With it, a user can look up specific information that he desires and find the appropriate report for his particular needs. However, the matrix is not completely cross-tabulated. For example, if a user is interested in wage information by race, and he finds a check by a particular publication for having wage information and some breakdowns by race, he cannot conclude that the report has wage information by race. All the matrix means is that some of the statistical information in the report is broken down by race, and that it also contains some wage information. The numbers in the matrix refer to the publication number in the first part of the paper. The final table entitled "labor supply" refers mostly to Office of Education material and the labor supply concept is primarily meant to be of some use in the calculation of future labor supply from the schools, including the vocational education schools.

8.1 Sources

1.0 National Science Foundation (NSF)

1.1 American Science Manpower - This was published biannually by NSF, and is based on a mail survey conducted by the same organization. The publication relates detailed characteristics, employment and earnings for American scientists. The most recent year available is 1970.

1.2 Employment of Scientists and Engineers in the U.S. - This was a one-time study done by NSF and the Bureau of Labor Statistics to establish historical series for employment of scientists and engineers. Appendix D of this publication is an excellent bibliography of data sources for employment of scientists and engineers.

2.0 Equal Employment Opportunities Commission (EEO)

2.1 EEO Report #2 Job Patterns for Minorities and Women In Private Industry Vol. 1 - This is published irregularly by EEO based on the EEO-1 form submitted by employers. Volume One gives employment of minority groups and women, for the nation and for the individual states. The most recent publication has data for 1970.

2.2 EEO Report Vol. 2 - This is the same as 2.1, except that it has the data for individual SMSA's in the United States.

3.0 Bureau of Labor Statistics (BLS)

3.1 Industry Wage Surveys - These are published irregularly for the various industries on the following page in three to five year cycles. They contain primarily industry wage data. The most recent surveys have data for 1972. The reports are based on personal interviews.

3.2 Employment and Earnings and the Monthly Report on the Labor Force - This is a monthly publication of BLS relating characteristics of the labor force, employment and unemployment, and earnings on a monthly basis. It draws from the Current Population Survey (CPS), BLS 790 data (Monthly Report on Employment Payroll and Hours), and BLS 1219 (Job Openings and Labor Turnover) as its primary sources.

3.3 Union Wage Reports (i.e. Union Wages and Hours in the ____ industry) - The publications for the various industries are usually released annually based on a mail survey of union leaders for selected building and printing trades, local transit and local trucking in 68 cities with a population of 100,000 or more. These reports relate wage and fringe information and changes in wage situations with some historical detail.

There is usually detailed city, and if relevant, trade detail. The most recent publications are based on data from 1972.

3.4 National Survey of Professional, Administrative, Technical and Clerical Pay - This publication provides nationwide salary averages and distribution for 80 work level categories covering 12 broad occupational groups. It is an annual publication based on a yearly survey covering mostly white collar workers. The most recent year has data for 1972.

3.5 The Handbook of Labor Statistics - This is an annual publication of BLS which relates employment, unemployment, earnings, and the general characteristics of the labor force. It is a summary of many of the BLS works, and uses such files as the CPS, Urban Employment Survey, BLS 790, BLS (or DL) 1219, industry and union wage surveys as well as information about prices and productivity. The most current edition is dated 1973, with data for 1972.

3.6 Area wage surveys - These publications report wage scales and wage movements for manufacturing and selected non-manufacturing industries in the selected SMSA's. These studies are done for 90 different labor market areas, usually at two year intervals for each SMSA. The most recent releases are for 1972.

3.7 Employment and Earnings Statistics for the United States - BLS issues this bulletin annually, and it summarizes employment and earnings data on a national basis. It gives earnings and hours data for production workers only. It has historical series for most of its data with the most recent release being for 1909-1972. However, most historical series do not go much further back than 1939.

3.8 Employment and Earnings States and Areas - This has much the same data as 3.7, except that it has data recorded separately for states and 210 separate areas (cities, SMSA's and for the individual burroughs of New York). Most of the historical data goes back only 20 years, although the most recent title is for 1939-1972.

3.9 Occupational Employment Statistics 1960-1970 - This is the latest in a series of BLS publications on occupational employment statistics. The first two (BLS report #305 and bulletin #1579) provide statistics for 1947-1966. This series is updated irregularly with the latest information being for 1970.

3.10 Scientific and Technical Personnel in Industry 1961-1969 - This report looks primarily at employment of scientific and technical personnel in private industry. It is the result of a mail survey that is conducted irregularly.

3.11 Tomorrow's Manpower Needs Vol. IV - "The National Industry-Occupational Matrix and Other Manpower Data" - This gives percentage figures or industry employment by occupation and occupational distribution by industry for 1960, and projected for 1975. The 1960 figures are based primarily on census data.

3.12 The Monthly Labor Review - This is a monthly publication, which in addition to carrying articles relevant to labor has a section entitled "current labor statistics", which contains household and payroll data on employment and earnings plus data on prices, productivity and work stoppages.

3.13 Indices of Output Per Manhour - Selected Industries 1939 & 1947-1972 - This publication develops a historical series for output per manhour for different mining and manufacturing industries, as well as rail and air transportation, and gas and electric utilities industries. This type of publication has been put out irregularly in the past.

3.14 Occupational Outlook Handbook 1974-75 edition - This is published every other year. It was first released in 1949 and has been revised several times. There is no statistical information in this publication, but it is a comprehensive volume examining different occupations, the nature of the work, training, employment outlook, and sources of other information.

3.15 Occupational Manpower and Training Needs - (BLS Bulletin #1701) This publication provides data on 1968 employment and projected 1980 requirements for selected occupations. When possible available training data is provided for each occupation.

4.0 The Manpower Administration (MA)

4.1 Area Trends in Employment and Unemployment - This is a monthly publication that details data on numbers in the work force, employment, and unemployment. As well, it singles out areas that have been hit by especially high unemployment rates.

4.2 Manpower Report of the President - This is an annual publication, and is an excellent collection of manpower information. It is two-thirds qualitative in nature, but it does have a detailed statistical section dealing with employment, unemployment, CPS data, and hours and earnings data for the nation and individual states and areas. It also has tables dealing with Manpower Development Training Act particpants, and some dealing with vocational education programs.

4.3 Dictionary of Occupation Titles - This is the 1965 or third edition of this basic source book for all counselors. The first edition came out in 1939 and the second in 1949. Volume I contains an alphabetical listing of occupations and descriptions. Volume II contains occupational categories, occupational group arrangement of titles and codes, worker trait arrangements of titles and codes, and an industry arrangement of titles as the main entries.

5.0 Office of Bureau Economics (OBE)

5.1 Business Statistics - 1973 - This is put out biannually by OBE as a supplement to the Survey of Current Business, a monthly publication of the same office. There is one section of this entitled "Labor Force, Employment, and Earnings". This section has historical data usually dating back to 1939 on most items. There is a section detailing the source materials for each table.

5.2 There are eight volumes of Growth Patterns in Employment by County, 1940-1950 and 1950-1960. These eight volumes deal with employment and changes in employment for the counties and States of the eight major regions of the United States as derived from the Census of Population. The change in employment for each county is shown with the amount by which it exceeds or falls short of the national average separated into industrial mix and regional share components. The influence of each of 32 industries on these employment changes is statistically detailed.

6.0 Bureau of the Census

6.1 Census of Government - Compendum of Public Employment - This is compiled every five years, with the latest edition being 1972. It has detailed data on employees and payrolls of federal, state, and local governments. It includes average monthly earnings of full-time employees.

6.2 City Employment in 1972 (Of governments) (Sample Employment

Data) - This has full-time employment, average full-time earnings, on a monthly basis for metropolitan areas as a whole. The latest year available is 1970.

6.3 Public Employment in 1972 (Sample Employment Data) - This contains public employment and payrolls of federal, state and local governments by function for each respective government. It has average earnings for state full-time employees, as well as some state and local detail.

6.4 Major Retail Centers in SMSA's - Census of Business-Retail Trade - This is compiled every five years by the Census Bureau. It has payroll information for the entire year, and numbers of paid employees for the week including March 12, by kind of business in the central business district. It has state, SMSA, city and central business district detail for cities over 100,000 in population. The most recent release is for 1972.

6.5 Selected Services-Area Statistics-Census of Business-Services - This is done every five years with the most recent being 1972. It has payroll information for the year, and paid employees for the week of March 12 (of 1966), by kind of business. It has state, SMSA, county and city detail.

6.6 Wholesale Trade-Area Statistics-Census of Business-Wholesale Trade - This information is also gathered every five years, with the most recent release being 1972. It has payroll data for 1971 and the first quarter of 1972, giving paid employees for the week of March 12, with state, SMSA, county, and city detail.

6.7 Census of Manufacturing-Area Statistics - This publication contains employment with industry detail, as well as aggregate manhours and wages for industries and areas.

6.8 The Decennial Census of the Population - This is the largest source of information on the population that one can find. Unfortunately, this information is only collected every ten years. In the Census are detailed demographic characteristics of the population as well as detailed wage and employment data. Any brief description would not do it justice. It should be consulted by all users of labor market information.

6.9 County and City Data Book, 1972 - This publication contains over 1000 pages of tables containing hundreds of data items on each county, SMSA, urbanized areas, unincorporated places of 25,000 or more, and cities of 25,000 or more population.

6.10 County Business Patterns - This annual report contains information on employment and payrolls by industry within counties, SMSA's and large cities.

7.0 U.S. Civil Service Commission

7.1 Current Federal Workforce Data - This is published biannually with the latest release being in January, 1970 based on the June, 1968 data. It is based on a 10% work history sample. It contains employment data for six month periods ending with December, 1967 and June, 1968. It covers a selected sample of 154 federal white collar occupations, representing about 95% of the total federal while collar workforce.

7.2 Federal Civilian Employment in the U.S. by Geographic Area - This is an annual publication. It contains employment by state, county, pay system, and selected agency, as well as SMSA detail.

7.3 Federal Workforce Outlook - This is in a series of annual publications that looks at projected federal workforce figures for a four year period. It has projected federal employment by occupation for 154 occupational series, which represents 95% of the total federal white collar workforce. The most recent report was issued in early 1974.

8.0 Office of Education (OE)

8.1 Digest of Educational Statistics - This is an annual publication of OE, which serves as an abstract of educational information. It is divided into five chapters: 1) All levels of education 2) Elementary and secondary education 3) Higher education 4) Federal programs of education, and 5) Selected statistics related to education. It deals with enrollment, teachers, income of schools, and of graduates of certain levels of education, and a multitude of other statistics. Two particularly interesting tables are Table 10, which presents occupations of employed persons by the level of school completed, sex and color, and Table 16, which presents total annual money income by years of school completed, sex, and age for persons over 25. The most recent year is 1973.

8.2 Vocational and Technical Education-Annual Report - This is the primary statistical work on vocational education. It is put out annually, and is based on the state reports received at the national offices. Most of the reporting is given with detail by state. The most recent report is for fiscal 1972.

8.3 Education and Training-A Chance to Advance - This report is put out annually by OE and HEW as a review of the year's activity under the Manpower Development And Training Act (MDTA). It is primarily qualitative in content, but it has a large statistical section with characteristics of the trainees, labor force status of those completing the program and some data on types of MDTA programs enrolled in. The most recent edition is for 1972.

8.4 Projections of Educational Statistics to 1982-83 - This is one in a series of publications that relates historical summary data for the ten previous years, and projections for the next ten years. It details enrollment, teachers, graduates, and expenditures for elementary, secondary, and higher education institutions. The most recent one was released in 1974. This is compiled by the National Center for Educational Statistics (NCES).

8.5 Students Enrolled for Advanced Degrees-Summary Data - This is an annual publication of OE based on a mail survey of universities. This part gives the enrollment in various courses for the whole country and enrollment by institution for all responding colleges. (This is for M.A. and Ph.D. candidates). The most recent release if for fall, 1970 (NCES).

8.6 Students Enrolled for Advanced Degrees-Institutional Data - This is much the same as 8.5, except that it details course enrollment for each university, as well as having sex and level of study detail.

8.7 Advance Statistics on Opening Fall Enrollment in Higher Education-Basic Information - This survey is carried out by mail questionnaire and lists higher education enrollment for each state for public and private institution by sex and student status. The most recent release is for fall, 1971 (NCES).

8.8 Directory-Public and Nonpublic Elementary and Secondary Day Schools Vols. I-V - This study is done by mail survey, and contains data for 1968-9. The first four volumes have information pertaining to public schools, one volume for each region of the country (i.e. Southwest), and a fifth volume for nonpublic schools. It contains enrollment data, numbers of teachers and number of graduates for 1968-9 for each reporting school.

8.9 Earned Degrees Conferred-part A - This is an annual publication. It lists graduates by level, school, sex, and has some aggregate data by state on individual fields of graduates. The latest edition covers 1970-1971 (NCES).

8.10 Earned Degrees Conferred-part B - This is much the same as 8.9 except that it presents the information for each individual school. For example, one can find how many M.A. degrees were conferred at a certain university in any of the listed fields.

8.11 Subject Offerings and Enrollments in Public Secondary Schools - This was a one time study done for the school year of 1961. There are hopes of a similar study in the future, perhaps in the next two years. It was based on a mail survey of 50% of the secondary schools in the United States. The main section lists, by state, the number of schools offering a particular course, and enrollment in each course including vocational education (NCES).

8.12 Preliminary Statistics of State School Systems - This is an annual publication that lists enrollment, instructors, expenditures, graduates, and money receipts from the government for each particular state (NCES).

8.13 Directory Public Schools in Large Districts with Enrollment and Staff by Race - This was a one time study one by mail survey for fall, 1967. The sample consisted of 10% of the schools in the country, which contained 70% of total enrollment. It presents enrollment and instructional staff by race for each school district. There are state and city names by each school district, so the data could be aggregated by state and city if the reader so desired (NCES).

8.14 Vocational Education and Occupations - This was put out in 1969 by OE to relate DOT occupation codes to types of vocational education programs. It is a reversible index. That is, it has DOT listings to vocational education programs in one section, and vocational education program to DOT code in the other.

9.0 Other Sources

9.1 Economic Report of the President and the Annual Report of the Council of Economic Advisors - This is put out annually by the federal government. The final section is a collection of statistical series. The general categories are: 1) National Income and Expenditure, 2) Population, employment, wages, and productivity, 3) Production and business activity, 4) Prices, 5) Money stock, credit, and finance, 6) Government finance, 7) Corporate profits and finance, 8) Agriculture, 9) International Statistics.

9.2 OBERS Projections, Regional Economic Activity in the U.S. - U.S. Water REsources Council. Volume I - Concepts, Volume II - OEA Economic Areas, Volumes III, IV - Water Resources Regions, Volume V - States - The volume projects employment and earnings from 1980 to 2020 with selected data for 1950, 1954, 1962, and 1969.

TABLE 3
DEMOGRAPHIC CHARACTERISTICS

	RACE	AGE	MARITAL STATUS	EDUCATIONAL ATTAINMENT	SEX
1.1		X		X	X
1.2					
2.1	X				X
2.2	X				X
3.1					
3.2	X	X	X		X
3.3					
3.4					
3.5	X	X	X	X	X
3.6					X
3.7					X
3.8					X
3.9					
3.10					
3.11					
3.12	X	X			X
3.15					
4.1					
4.2	X	X	X	X	X
5.1					
5.2					
6.1					
6.2					
6.3					
6.4					
6.5					
6.6					
6.7					
6.8	X	X	X	X	X
6.9	X	X	X	X	X
6.10					
7.1					
7.2					
7.3					
8.1	X	X			X
8.2					X
8.3	X	X		X	X
8.4				X	X
8.5					X
8.6					X
8.7					X
8.8					

TABLE 3 (Continued)

DEMOGRAPHIC CHARACTERISTICS

	RACE	AGE	MARITAL STATUS	EDUCATIONAL ATTAINMENT	SEX
8.9				X	
8.10				X	X
8.11					
8.12					
8.13	X				
9.1	X	X			X
9.2					

TABLE 4

EMPLOYMENT

	EMPLOYMENT DATA	OCCUPATION DETAIL	INDUSTRY DETAIL	UNEMPLOY. DATA	HOURS OF WORK	PROJECT-IONS
1.1	X	X	X	X		
1.2	X	X	X			
2.1	X	X	X			
2.2	X	X	X			
3.1					X	
3.2	X	X	X	X	X	
3.3					X	
3.4	X	X			X	
3.5	X	X	X	X	X	
3.6						
3.7	X		X		X	
3.8	X		X		X	
3.9	X	X	X			
3.10	X	X	X			
3.11	X	X	X			X
3.12	X	X			X	
3.15	X	X				X
4.1	X			X		
4.2	X	X		X		X
5.1	X		X	X	X	
5.2	X		X			X
6.8	X	X	X	X		
6.9	X	X	X	X		
6.10	X		X			
7.1						
7.2	X	X				
7.3	X	X				
7.4	X					
7.5	X		X			
7.6	X					
7.7	X		X		X	
8.1	X	X				
8.2	X		X			
8.3	X	X				X
9.1	X		X	X		
9.2	X		X			X

TABLE 5

WAGES

	WAGE AND SALARY DATA	INDUSTRY DETAIL	OCCUPATIONAL DETAIL
1.1	X	X	X
1.2			
2.1			
2.2			
3.1	X		X
3.2	X		X
3.3	X		X
3.4	X	X	X
3.5	X	X	X
3.6	X	X	X
3.7	X	X	
3.8	X	X	
3.9			
3.10			
3.11			
3.12	X	X	
3.15			
4.1			
4.2	X		
5.1	X	X	
5.2	X	X	
6.8	X	X	X
6.9	X	X	
6.10	X	X	
7.1	X		
7.2	X		X
7.3	X		X
7.4	X	X	
7.5	X	X	
7.6	X		
7.7	X	X	
8.1			
8.2			
8.3			
9.1	X	X	
9.2	X	X	

TABLE 6

GEOGRAPHIC BREAKDOWN

	NATIONAL	STATE	SMSA	INSTITUTIONAL
1.1	X	X	X	
1.2	X			
2.1	X	X		
2.2			X	
3.1	X		X	
3.2	X	X	X	
3.3	X	X[1]	X	
3.4	X			
3.5	X	X	X	
3.6			X	
3.7	X			
3.8		X	X	
3.9	X			
3.10	X	X		
3.11	X			
3.12	X			
3.15				
4.1	X		X	
4.2	X	X	X	
5.1	X			
5.2			X[2]	
6.1	X	X	X[3]	
6.2			X	
6.3	X	X	X	
6.4	X	X	X	
6.5	X	X	X	
6.6	X	X	X	
6.7	X	X	X	
6.8	X	X	X	
6.9	X	X	X	
6.10	X	X	X	
7.1	X			
7.2	X	X	X	
7.3	X			
8.1	X	X		
8.2	X	X		
8.3	X	X		
8.4	X	X		
8.5	X	X		
8.6				X
8.7	X	X		

TABLE 6 (Continued)

GEOGRAPHIC BREAKDOWN

	NATIONAL	STATE	SMSA	INSTITUTIONAL
8.8				X
8.9	X	X		
8.10				X
8.11	X	X		
8.12	X	X		
8.13				X
9.1	X			
9.2			X[2]	

[1] by region

[2] by OBE area

[3] by county

TABLE 7

LABOR SUPPLY

	SCHOOL ENROLLMENT	SCHOOL INSTRUCTORS	SCHOOL COMPLETED	FOLLOW-UP STUDIES	COURSE ENROLLMENT	COURSE COMPLETIONS	PROJECTIONS
4.2	X		X	X			
8.1	X	X	X		X		
8.2	XD	X			X		
8.3	X			X	X		
8.4	X	X	X				X
8.5	X				X		
8.6	X				X		
8.7	X						
8.8	X	X	X				
8.9			X			X	
8.10			X			X	
8.11	X				X		
8.12	X	X	X				
8.13	X	X					

APPENDIX M

MINORITY EMPLOYMENT DATA SOURCES FOR SMSA'S

Malcolm S. Cohen and Nira Shamai

9.0 Introduction

The United States Department of Labor's Office of Federal Contract Compliance requires Federal contractors to do an analysis of minority worker utilization in its major job categories. The employer is required to compare his minority employment with the available labor supply in his area. For 1970 it is possible to make a comparison between employment in his establishment and employment from the 1970 Census. For non-Census years such a comparison is more difficult. The BLS 790 Program does provide some breakouts of employment by sex and industry at the national level but not by occupation. The Occupational Employment Survey provides occupational-sex breakouts but it is not yet available for all areas, nor are race breakouts available. The Equal Employment Opportunity Commission (EEOC) collects data on minority employment; however, it is not comparable to Census information for a number of reasons:

1) EEOC information is available for only nine broad occupational classifications while Census information is available for 297 detailed occupations.

2) EEOC surveys firms while the Census interviews households.

3) EEOC eliminates companies which are smaller than 100 employees and that do not have government contracts over $10,000. The Census ascertains employment status for everyone over age 14.

4) EEOC counts jobs. The Census counts persons. Since a person may have more than one job he can be counted at each employer where he works.

5) EEOC covers establishments located in the area. Census covers persons living in the area regardless of where they work.

6) The EEOC survey of establishments excludes persons not working in establishments such as self-employed persons. EEOC also excludes public administration. However, special Census tabulations are presented excluding public administration.

7) The Census applies to the Census week in April, 1970. The EEOC

applies to any pay period during December, 1969 - April, 1970.

Using the 1970 Census as a benchmark we investigated updating the racial and sex distributions of employment by occupation and industry using EEOC data. The first job in such an analysis is a reconciliation of differences between EEOC data and Census data for 1970. This could serve as a basis for a synthetic data base. Racial distributions in the Census and EEO-1 can be applied to industries serveyed by the Occupational Employment Survey. For the states not participating in the OES program the data might be used to provide the distribution of employment by sex and race across broad occupational groups for labor market areas.

The analysis was undertaken for three SMSA's for 1970: Denver, Detroit, and Milwaukee. Census data is based primarily on a 2% Public Use Sample; however, some tabulations are based on the published 20% sample and others are based on only a 1% sample.

9.1 Milwaukee

Table 8 presents the distribution by sex by occupation of the total employed in the Milwaukee SMSA for 1970. A Public Use Sample tape was purchased from the Census Bureau and some of the tabulations shown are based on tabulations made from the Public Use Sample using the MICRO retrieval program. A comparison of the results obtained from the Public Use Sample and published Census data is shown in Table 8. The published information was not available, however, for occupation by industry by race. While this tabulation is not too important for Milwaukee, it was of much greater importance for Detroit.

Table 9 presents a comparison of the percent female and percent Negro derived from EEOC and the 1970 Census. Other minority information for Indians, Orientals and Spanish-Americans is available, but tabulations from a 2% sample of Census records are not too meaningful for groups this small.

Because of the many differences between EEOC and Census it would be surprising if many of the entries in Table 9 were very close to one another. One measure of closeness is whether or not the EEOC estimates are within two standard errors of the Census estimates. By this criterion quite a few of the estimates are not significantly different from one another. Even in cases where the two are significantly different by statistical criteria, the differences are sometimes economically meaningless. For example, 25.6% of the operatives are female according to EEOC estimates and 27.7% are female according to Census estimates. It is inconceivable that manpower planners would be led astray if they assumed 26.5% of the operatives were female in Milwaukee. A more serious concern is whether the occupational classification, operative, is a meaningful one for manpower policy.

The most serious discrepancy between EEOC and Census in the percent female is for laborers. This discrepancy held for all three SMSA's analyzed as tables for the other SMSA's will show. An analysis of this difference was carried out. We believe it is due to differences in reporting by households and firms. Women are reluctant to report themselves as laborers, whereas employers according to EEO-1 instructions are asked to report on the number of "laborers (unskilled)" workers. In the Census women might classify themselves as operatives (semi-skilled), service workers or clerical workers to raise their self image. This explanation is consistent with our findings that only 1.2% of all Milwaukee women workers classified themselves as laborers according to the 1970 Census while 7.8% of all women workers were classified as laborers according to EEOC reports.

The tendency of the Census to overstate the percent of female officials, managers, professionals and technicians is explainable using a similar agrument. EEOC reports 11% of all females to be in managerial, professional and technical occupations compared with 17.5% in the Census. If the tendency of women is to upgrade their occupational status one would expect more women to report themselves as officials, managers, professionals and technicians than employers report. An equally plausible explanation for both results is that employers downgrade the reported skill level of women. Unless occupational classifications are based on fairly well-defined skill definitions uniformly applied it would be difficult to choose between the competing explanations.

Another interesting result of the comparison is that there is a higher proportion of females reported employed in the Census in every occupation except sales workers and laborers. This is probably due to the cutoff which removes most firms with an employment of less than

100 workers from the EEOC universe. This would suggest women are more likely to work in smaller establishments than men. It could also result in part from the dramatic difference in the distribution of laborers between the two surveys.

Table 10 compares the total number of jobs and the total number of employees as well as an estimate of the number of units with less than 100 employees based on Treasury Form 941 reports. The EEOC file contains some units with an employment of less than 100 workers. This includes certain Federal contractors as well as companies with 100 or more employees with some locations having less than 100 workers. For example in manufacturing according to EEOC reports 4,788 workers were employed in establishments with less than 100 workers. Thus approximately 165,000 persons were reported to be employed in manufacturing by EEOC in firms with 100 or more employees and about 40,000 were estimated to be in firms of less than 100 from Treasury Form 941 reports. This is almost exactly the same as employment reported by the Census, but less than total employment reported by Treasury Form 941. The differences could easily be explained by differences in pay period reported or differences in the geographic area that the firm reports. Establishment employment for March, 1970 reported for manufacturing from the BLS-790 program was between these estimates -- 211,700 [16].

The industry with the largest discrepency between EEOC coverage and Census coverage is construction. The employment reported in the Census is eight times that reported in EEOC. As one might expect, this can be explained in large part from the many small construction firms. Seventy-five percent of the construction firms in Milwaukee have less than 100 employees.

Table 11 presents a count of employment by occupation and industry in EEOC and Census. Table 12 presents a breakout of percent female and percent Negro for the same industry-occupation matrix. Table 12 could be disaggregated by the nine occupations shown in Table 8 for use for affirmative action uses. However, for comparison to the Census we had to aggregate in this manner to minimize the number of statistically insignificant cells.

9.2 Denver and Detroit

Tables 10, 11 and 12 are repeated for Denver and Detroit. The discussion for Milwaukee applies equally well to these SMSA's.

Of special interest in the Detroit SMSA is Table 16 which compares the percent female and percent Negro by occupation. The findings for females is similar to that reported in Milwaukee. The level of Negro employment in Detroit is far greater that in either of the other cities. Therefore, it is of greater interest to compare percent Negro in EEOC and Census for Detroit. The same phenomenon we observed for females holds for blacks. Census data overstates the percent black officials, managers, professionals and craftsmen and understates the laborers, service workers and operatives. The overstatement is greatest for skilled workers and the understatement is greatest for unskilled workers.

Another test we made was to compare Social Security, Census and EEOC records to see how accurate our overall proportion of black was:

Detroit, SMSA, 1970

	Persons Covered by Social Security % Nonwhite	Census % Negro	Jobs EEOC % Negro
Manufacturing	16.6%	16.5%	20.2%
Non-Manufacturing (Excluding agriculture and government)	13.4%	14.6%	17.9%

The Social Security data does not separate Negroes from other nonwhites and about 1% of the employed in Detroit are "other nonwhites." Thus the percent Negro from Social Security records would be about 15.5% for manufacturing and 12.5% for non-manufacturing. Social Security records refer to all persons employed during any time during first quarter, 1970 classified by industry of major income.

Differences in multiple job holding between whites and blacks account for a little more of the difference. A special tabulation was made using Social Security data comparing persons counted once for every job they held during the first quarter, 1970 and once only at the major job they held. For example, there were 104,200 nonwhites whose major job was in manufacturing and 111,400 whose major job was in non-manufacturing. These 215,600 blacks held 110,100 jobs in manufacturing and 136,600 jobs in non-manufacturing. However, recomputing percent nonwhite on the basis of jobs rather than persons makes a difference of only 0.2 percent.[1]

Thus it appears that after all of the adjustments have been made the difference of about four percentage points between EEOC and Social Security is due to the exclusion by EEOC of small firms and non-reporting by larger firms. Firms with few blacks are probably more likely not to report in EEOC. However, non-reporting for Social Security is much more difficult, especially in manufacturing.

9.3 Previous Tabulations

An example of a tabulation from EEOC reports is shown in Table 19. Such tabulations are available for major SMSA's. The tabulation shown is for Denver for 1967. While the detail presented in the report is certainly useful and should be continued, our tables make analysis and comparison to the Census easier.

9.4 Conclusions and Recommendations

In addition to pointing out some insights into differences between EEOC and Census data as sources of Equal Employment Opportunity information, the study illustrated the value of a computer language like MICRO for analysis of survey data. On a number of occasions it was desirable to retabulate data in an unforeseen form as hypotheses were suggested. This was readily possible using MICRO. The version of MICRO used for this study lacked a few facilities that would have made the analysis even more economical and easy to do. These facilities included the ability to recode data fields such as occupation and industry. Also some "bugs" in the cross tabulation facilities of MICRO caused much grief to the analyst. These limitations were removed in the current version of MICRO but not in time for use in this study.

---

[1] The multiple job adjustment takes account of differences due to changing jobs during the quarter as well as holding two jobs at one time during the quarter. However, a remaining difference between EEOC and Social Security is that EEOC refers to a payroll period while Social Security refers to the quarter. Thus, some unemployed persons not reported in EEOC may show up in Social Security records if they worked at some times in the quarter. The similarity of Census to Social Security records in manufacturing combined with the limited amount of commuting outside of the SMSA suggests differences in the duration of employment account for very little of the remaining difference.

Another by-product of the study is some suggested tabulations which could be run for large SMSA's in the United States using both 1970 Census data and EEOC data. Further disaggregation for races other than Negro and disaggregation for the nine occupation groups by summary industry groups would also be desirable in addition to the basic tabulations shown in Tables 9, 11 and 12. Such tabulations are available in sixth count Census tapes and should be carried out by EEOC to aid manpower planners concerned with affirmative action programs for more current information. The manpower planners may find our study useful in understanding differences between the two information sources.

DOCUMENTATION OF INDUSTRY, OCCUPATION CODES

A. The matching Census Occupations to fit the EEO Occupations group.

EEO Occupation Groups	Census Recoding Occupation
Officials and Managers	07
Professional	01, 02, 04, 06
Technicians	03, 05
Sales	08
Office and Clerical + (White Collar) Trainees	09
Craftsmen	10, 11
Operatives + On the job trainees Production	12, 13
Laborers	14
Service Workers	16, 17, 18, 19, 20

B. Matching Census Industries to fit the groups of SIC industry code of EEOC Survey (EEO-1).

Industry Groups	Census INDUSTRY Codes	EEO-1 SIC Groups Code
Manufacturing		
Durables		
Metal	05, 06	33-34
Machinery	07, 08	35-36
Transportation	09	37
Other Durables	04, 10	24-25, 19, 32, 38, 39
Non-Durables		
Food	11	20-21
Textiles, Printing Other Non-Durables	12, 13, 14, 15	22-23, 26-31
Non-Manufacturing		
Mining	02	10-14
Construction	03	15-17
Transportation and Communication	16, 17, 18 19, 20	40-41, 42-49
Wholesale	21	50
Retail	22, 23, 24, 25, 26	52-59
Finance, Insurance Real Estate	27, 28	60-67
Services, Non-Profit	29, 30, 32, 33, 34 35, 36, 37, 38, 39	70, 72-73, 75-76 78-82, 84, 86, 89

TABLE 8

Milwaukee SMSA, 1970

Comparison Between 20% Sample of the Census and 2% Sample

Total Employed (non-farm excluding private household workers)[1]

	Published Census			Census 2%		
Occupation	Total	Male	Female	Total	Male	Female
Official & Managers	41450	35502	5948	40400	34500	5900
Professionals & Technicians	85613	52622	32991	88050	55350	32700
Sales Workers	44539	24518	20021	47600	27150	20450
Office & Clerical	108240	27000	81240	108350	26450	81900
Craftsmen	79379	75216	4163	79300	74800	4500
Operatives	113854	82568	31286	118150	85650	32500
Laborers	23188	20417	2771	22200	19350	2850
Service Workers	68089	28215	39874	72050	28850	43200

Source: U.S. Census, 1970 Census of Population, Table 180, Detailed Characteristics, PC(1) - D51 and public use sample data tapes.

[1] Public administration is included in this table but excluded from future tables because government workers are not covered by EEO-1 reports which are compared in later tables to census tables.

TABLE 9

Milwaukee SMSA, 1970
Comparison Between EEOC and Census 2% Sample[1]

Occupation	Total Employed		% of Female		% of Negro	
	EEOC	Census	EEOC	Census	EEOC	Census
			(1) Total			
Officials & Managers	23874	38150	7.8	14.6	1.1*	0.8
Professionals & Technicians	33977	85000	25.8	37.8	2.0*	2.9
Sales Workers	27105	47600	48.6	43.0	2.1*	2.4
Office & Clerical	41954	99500	78.7*	78.0	3.9*	4.2
Craftsmen	40206	77200	3.7	5.8	3.6*	4.4
Operatives	72153	117200	25.8	27.7	12.2	10.4
Laborers	23610	21050	31.9	12.8	14.1*	14.3
Service Workers	20213	65500	60.5	64.6	15.2	11.5
			(2) Manufacturing			
Officials & Managers	15056	10150	2.2*	4.4	0.8*	---
Professionals & Technicians	16470	22450	5.8*	6.9	1.0*	1.8
Sales Workers	5468	7300	6.7*	6.2	1.8	2.1
Office & Clerical	19381	28000	69.7*	65.9	2.3*	2.3
Craftsmen	32716	39600	3.1	6.3	3.9	5.9
Operatives	59884	86050	26.6	29.7	13.0	10.0
Laborers	18356	7150	36.4	14.0	15.5*	15.4
Service Workers	2973	4050	19.3*	19.8	11.5*	11.1
			(3) Non-Manufacturing			
Officials & Managers	8818	28000	17.6*	18.3	1.7*	1.1
Professionals & Technicians	17507	62550	44.8	48.9	2.9*	3.3
Sales Workers	21637	40300	59.2	49.6	2.1*	2.5
Office & Clerical	22573	71500	86.4	82.8	5.2*	4.9
Craftsmen	7490	37600	6.2*	5.2	2.4*	2.7
Operatives	12269	31150	22.2*	22.1	8.5*	8.9
Laborers	5254	13900	16.4*	12.2	9.4*	13.7
Service Workers	17240	61450	67.6*	67.6	15.8	11.5

---

[1] Excluding public administration and private household. Census 2% sample excludes agriculture and EEOC figures include agriculture.

*Estimate from EEOC within two standard errors of Census estimate.

TABLE 10

Milwaukee SMSA, 1970
Analysis of Employment by Size of Firm by Industry

	Total Employment, 1970			County Business Patterns, 1970				
Industry	EEOC (Jobs) (1)	Census (Employees) (2)	Percent EEOC of the Census (3)= (1)/(2)	Total Employment (4)	Total Reporting Units (5)	Total No. of Units With Less Than 100 Employees (6)	Total No. of Employee's in Units With Less Than 100 (7)	% of Employment in Units With Less Than 100 (8)=(7)/(4)
Construction	3,012	24,950	12.1%	19,960	2,052	2,032	15,000	75%
Manufacturing	170,304	204,750	83.2%	219,585	2,472	2,132	40,000	20%
Transportation & other Public Util.	17,544	31,050	56.5%	27,966	652	616	16,000	57%
Wholesale Trade	8,309	24,800	33.5%	31,759	2,181	2,145	22,000	70%
Retail Trade	34,919	95,750	36.5%	91,242	6,118	6,039	64,000	70%
Finance	12,365	30,900	40.0%	28,917	2,194	2,155	17,000	58%
Services	36,639[1]	138,600	26.4%	85,600	6,677	6,554	46,000	54%

[1] Includes services, mining, and agriculture due to confidentiality restrictions.

Source: U.S. Department of Commerce, County Business Patterns, Wisconsin CBP-70-51, Table 3, p. 120, 1971.

TABLE 11

Milwaukee SMSA, 1970
Employment Industry by Occupation
Comparison Between EEOC and Census 2% Sample[1]

Industry	Occupation					
	Prof. Tech. Mgrs.		Sales Service, Clerk		Craft. Oper. Laborer	
	EEOC	Census	EEOC	Census	EEOC	Census
			Manufacturing			
Durables						
Metal	4310	4750	3908	5400	21283	27450
Machinery	15719	16900	12558	14650	50140	53550
Transportation	4451	1100	2469	1750	12735	9350
Other durables	1900	2550	1408	3900	5459	12850
Non-Durables						
Food	2278	1750	2882	3100	8546	9050
Textiles, Printing, other Non-Durables	2868	5550	4597	10550	12793	20550
			Non-Manufacturing			
Construction	223	3050	189	2400	2600	19500
Transport. Comm.	3422	4000	4284	7850	9838	19200
Wholesale	2000	3600	3459	11150	2850	10050
Retail	3477	12200	25568	64550	5874	19000
Finance Ins. Real Estate	3445	6700	8771	23350	149	850
Service Non profit[2]	13758	60900	19179	63800	3702	13900

---

[1] Exluding public administration and private household. Census 2% sample excludes also agriculture and mining.

[2] EEOC figures include mining and agriculture due to confidentiality restrictions.

TABLE 12

Milwaukee SMSA, 1970
Comparison Between Census 2% and EEOC Industry by Occupation
Percent Female and Percent Negro[1]

	Occupation											
	Prof. Tech. Mgrs.				Sales, Service, Clerk				Craft. Oper. Laborer			
	% Female		% Negro		% Female		% Negro		% Female		% Negro	
Industry	EEOC	Census	EEOC	Census	EEOC	Census	EEOC	Census	EEOC	Census	EEOC	Census
					Manufacturing							
Durables												
Metal	3.6*	5.3	0.1*	1.1	55.6*	46.3	2.8*	2.8	16.0*	16.0	12.1*	11.7
Machinery	3.0*	3.6	0.6*	0.9	55.2*	57.7	2.4*	2.4	22.1*	20.6	8.5	6.0
Transportation	1.6*	--	1.3*	--	43.9*	37.1	3.7*	8.6	8.7*	7.5	17.6*	18.2
Other Durables	3.8*	5.9	5.8*	3.9	62.6*	51.3	1.8*	5.1	43.8	28.0	7.4*	17.1
Non-Durables												
Food	6.9*	14.3	2.1*	2.9	43.1	58.1	5.9	1.6	8.7*	13.3	9.0*	7.7
Textiles, Printing, other Non-Durables	10.1*	13.5	1.3*	0.9	46.4*	40.8	4.0*	3.3	38.5*	39.4	12.2	9.5

(continued)

TABLE 12

Milwaukee SMSA, 1970
Comparison Between Census 2% and EEOC Industry by Occupation
Percent Female and Percent Negro[1]

	Occupation											
	Prof. Tech. Mgrs.				Sales, Service, Clerk				Craft. Oper. Laborer			
	% Female		% Negro		% Female		% Negro		% Female		% Negro	
Industry	EEOC	Census	EEOC	Census	EEOC	Census	EEOC	Census	EEOC	Census	EEOC	Census
					Non Manufacturing							
Construction	3.0*	3.3	0.4*	--	43.9*	45.8	1.0*	--	0.0	1.5	3.1*	5.1
Transport. Comm.	9.1	21.3	1.3*	3.8	60.0*	59.2	5.1*	5.1	0.8	3.4	3.5	6.5
Wholesale	4.2*	5.6	2.2*	--	42.8*	39.5	2.8	1.3	17.1	12.4	8.2*	8.0
Retail	26.5*	21.7	1.9*	0.8	73.9	70.6	3.8*	4.8	35.5	18.4	7.6*	5.5
Finance, Ins. Real Estate	15.9*	23.1	1.0*	--	72.8	63.0	6.2	3.4	47.6*	35.3	4.0*	--
Service Non Profit[2]	54.5	50.0	3.3*	3.4	75.8	78.8	13.1*	11.2	35.7*	32.4	16.3	10.8

---

[1] Excluding public administration, private household. Census 2% sample excludes also agriculture and mining.

[2] EEOC figures include mining and agriculture due to confidentiality restrictions.

*Estimate from EEOC within two standard errors of Census estimate.

TABLE 13

Denver SMSA, 1970
Comparison Between EEOC and Census 2% Sample[1]

	Total Employed		% of Female		% of Negro	
Occupation	EEOC	Census	EEOC	Census	EEOC	Census
			(1) Total			
Officials & Managers	17400	44500	11.8	15.1	0.9*	1.2
Professionals & Technicians	31832	88000	24.1	38.3	1.7*	1.4
Sales Workers	17851	41900	35.6*	34.4	1.8*	1.2
Office & Clerical	35320	93300	77.9*	78.5	3.4*	3.2
Craftsmen	21512	56100	3.9*	5.0	2.1*	1.7
Operatives	25175	58250	19.9	25.1	4.5*	3.8
Laborers	12203	19150	20.8	8.1	6.8*	6.0
Service Workers	16833	54850	52.9	59.2	14.9	8.4
			(2) Manufacturing			
Officials & Managers	5478	6850	2.0	10.2	0.7*	0.7
Professionals & Technicians	11849	14500	4.0	8.4	1.3*	0.7
Sales Workers	2148	4700	5.4*	6.4	1.1*	4.3
Office & Clerical	7360	12650	69.4*	69.6	3.9*	2.4
Craftsmen	8854	16600	4.9*	4.7	2.8*	1.9
Operatives	12164	26400	25.7	34.9	5.0*	4.1
Laborers	5629	2950	31.3	1.7	7.1*	5.1
Service Workers	1258	1200	11.2*	12.5	18.4*	12.1
			(3) Non-Manufacturing			
Officials & Managers	11922	37650	16.3*	15.9	1.0*	1.3
Professionals & Technicians	19983	73500	36.0	43.8	2.0*	1.5
Sales Workers	15703	37200	39.8*	37.9	1.9	.8
Office & Clerical	27960	80650	80.2*	79.9	3.3*	3.3
Craftsmen	12658	39500	3.3	5.1	1.7*	1.5
Operatives	13011	31850	14.4*	17.3	4.1*	3.5
Laborers	6574	16200	11.9*	9.3	6.7*	6.2
Service Workers	15575	53650	56.3	60.2	14.7	8.3

---

[1] Excludes public administration and private household. Census 2% sample excludes agriculture and EEOC figures include agriculture.

*Estimate from EEOC within two standard errors [illegible]ensus estimate.

TABLE 14

Denver SMSA, 1970
Employment Industry by Occupation
Comparison Between EEOC and Census 2% Sample[1]

Industry	Occupation					
	Prof. Tech. Mgrs.		Sales Service, Clerk		Craft. Oper. Laborer	
	EEOC	Census	EEOC	Census	EEOC	Census
Manufacturing						
Durables						
Metal	766	1700	571	1350	3654	5650
Machinery	5933	5050	3272	3250	6347	7100
Transportation	5039	4050	1184	1000	1934	3250
Other Durables	2159	4050	1266	2650	5151	9000
Non-Durables						
Food	1146	1250	1692	1100	5175	6950
Textiles, Printing, Other Non-Durables	2260	5250	2767	9200	4215	1400
Non-Manufacturing						
Construction	533	5350	456	3050	4611	20800
Transport. Comm.	6644	8050	10687	13650	13391	19500
Wholesale	2666	6100	4263	14350	4309	10550
Retail	3307	14750	17897	49850	5235	20950
Finance, Ins. Real Estate	3405	7500	8628	24200	306	800
Service Non Profit[2]	14398	66700	16870	65000	3926	14300

---

[1] Excluding public administration and private household. Census 2% sample excludes also agriculture. Mining not shown due to possible disclosure.

[2] EEOC figures include agriculture.

TABLE 15

Denver SMSA, 1970
Comparison Between Census 2% and EEOC Industry by Occupation
Percent Female and Percent Negro[1]

	Occupation											
	Prof. Tech. Mgrs.				Sales, Service, Clerk				Craft. Oper. Laborer			
	% Female		% Negro		% Female		% Negro		% Female		% Negro	
Industry	EEOC	Census	EEOC	Census	EEOC	Census	EEOC	Census	EEOC	Census	EEOC	Census
					Manufacturing							
Durables												
Metal	2.1*	8.8	0.9*	---	48.1*	59.3	3.8*	3.7	9.9*	11.5	6.7	2.7
Machinery	1.8*	6.9	1.8*	---	46.5*	60.0	6.1*	3.1	13.0	25.4	4.9*	4.2
Transportation	2.5*	4.9	0.3*	---	61.7*	55.0	4.8*	---	8.0*	13.8	3.1*	1.5
Other Durables	6.8*	3.7	0.8*	1.2	68.1	41.5	1.5*	1.9	33.3	17.8	3.0*	3.3
Non-Durables												
Food	4.3*	12.0	0.6*	---	31.2*	27.3	4.1*	9.1	22.4*	16.5	6.2*	4.3
Textiles, Printing, Other Non-Durables	6.1	19.0	1.3*	1.9	52.0*	49.5	6.1*	4.3	24.6*	31.4	3.7*	3.2

(continued)

TABLE 15

Denver SMSA, 1970
Comparison Between Census 2% and EEOC Industry by Occupation
Percent Female and Percent Negro 1

	Occupation											
	Prof. Tech. Mgrs.				Sales, Service, Clerk				Craft. Oper. Laborer			
	% Female		% Negro		% Female		% Negro		% Female		% Negro	
Industry	EEOC	Census	EEOC	Census	EEOC	Census	EEOC	Census	EEOC	Census	EEOC	Census
	Non Manufacturing											
Construction	4.6*	2.8	1.5*	--	49.7*	49.2	1.5*	6.6	1.9	0.7	2.0*	2.4
Transport. Comm.	10.0*	9.9	0.6*	0.6	62.6*	65.6	5.8	3.3	0.6	4.1	3.1*	3.1
Wholesale	3.2*	5.7	1.1*	--	38.7*	40.4	2.1*	2.8	8.1	12.3	2.4*	2.8
Retail	23.4*	18.6	1.5*	1.7	59.4	63.7	3.7	2.4	20.1	15.5	3.9*	3.1
Finance, Ins. Real Estate	11.0	24.7	0.9*	1.3	71.9	54.8	3.3*	2.7	14.0*	31.3	5.2*	--
Service Non Profit 2	50.2*	48.3	2.3*	1.9	69.3	74.5	10.8	6.8	38.3	21.7	8.0	4.9

[1] Excluding public administration and private household. Census 2% sample also excludes agriculture. Mining not shown due to possible disclosure.

[2] EEOC figures include agriculture.

*Estimate from EEOC within two standard errors of Census estimate.

TABLE 16

Detroit SMSA, 1970
Comparison Between EEOC and Census 2% Sample[1]

Occupation	Total Employed		% of Female		% of Negro	
	EEOC	Census	EEOC	Census	EEOC	Census
(1) Total						
Officials & Managers	71121	106050	9.3	15.7	3.9	5.0
Professionals & Technicians	100310	230800	19.1	35.6	6.4	8.2
Sales Workers	61507	117100	48.7	41.8	7.5*	6.5
Office & Clerical	127138	265600	73.4	74.2	15.2	12.8
Craftsmen	100706	233750	2.0	3.2	7.4	10.4
Operatives	216203	312900	12.8	19.8	32.1	26.1
Laborers	52215	61100	21.1	7.9	36.0	23.1
Service Workers	61199	169150	43.8	60.4	36.6	26.1
(2) Manufacturing						
Officials & Managers	37140	25600	1.7	5.5	3.6	1.2
Professionals & Technicians	44272	70850	3.6*	4.6	2.0*	2.7
Sales Workers	8189	18050	11.6	8.6	3.7*	3.0
Office & Clerical	40234	73300	56.8*	56.1	8.3*	7.1
Craftsmen	64808	135900	1.4	2.3	6.4	10.2
Operatives	177134	231150	13.0	19.6	34.5	28.1
Laborers	28230	20950	23.4	7.9	29.6	26.7
Service Workers	12242	18450	16.2*	14.6	33.7*	30.9
(3) Non-Manufacturing						
Officials & Managers	33981	80450	17.6	18.9	4.3	6.3
Professionals & Technicians	56038	159950	31.5	49.1	9.9*	10.6
Sales Workers	53318	99050	54.5	47.9	8.1*	7.2
Office & Clerical	86904	192330	81.1*	81.0	18.3	14.9
Craftsmen	35898	97850	3.3	4.4	9.5*	10.7
Operatives	39069	81750	12.1	20.2	21.3*	20.5
Laborers	23985	40150	18.4	7.9	43.6	21.2
Service Workers	48957	150700	50.7	66.0	37.3	25.5

---

[1] Excludes agriculture, public administration and private household.

*Estimate from EEOC within two standard errors of Census estimate.

TABLE 17

Detroit SMSA, 1970
Employment Industry by Occupation
Comparison Between EEOC and Census 2% Sample[1]

Industry	Occupation					
	Prof. Tech. Mgrs.		Sales Service, Clerk		Craft. Oper. Laborer	
	EEOC	Census	EEOC	Census	EEOC	Census
Manufacturing						
Durables						
Metal	6904	9000	5392	12900	40347	57500
Machinery	17584	18500	12748	16050	64337	64950
Transportation	45621	47050	28729	45550	130587	181250
Other Durables	2700	7450	1928	9100	7656	35800
Non-Durables						
Food	1250	1600	3136	2750	7124	11900
Textiles, Printing Other Non-Durables	7353	12850	8732	23450	20121	36600
Non-Manufacturing						
Construction	1038	9250	794	7750	6323	49500
Transport. Comm.	11809	13000	22854	28900	32409	45900
Wholesale	8739	11500	12721	28750	17169	23600
Retail	11999	34400	66961	157800	16103	57850
Finance, Ins. Real Estate	11150	15900	26162	55300	458	2350
Service Non Profit	45100	156100	59373	162950	25917	39700

---

[1] Excluding agriculture, public administration, and private household. Mining not shown due to possible disclosure.

TABLE 18

Detroit SMSA, 1970
Comparison Between Census 2% and EEOC Industry by Occupation
Percent Female and Percent Negro[1]

	Occupation											
	Prof. Tech. Mgrs.				Sales, Service, Clerk				Craft. Oper. Laborer			
	% Female		% Negro		% Female		% Negro		% Female		% Negro	
Industry	EEOC	Census	EEOC	Census	EEOC	Census	EEOC	Census	EEOC	Census	EEOC	Census
					Manufacturing							
Durables												
Metal	2.4*	5.0	3.3*	1.7	39.4*	43.8	10.9	7.0	13.2	11.1	22.0	14.4
Machinery	2.2*	5.1	3.2	1.9	43.2*	48.0	14.8	3.4	13.3*	11.6	22.1	6.5
Transportation	2.0*	2.7	2.5*	2.3	41.3*	40.7	13.7	13.9	6.3	8.9	32.4	29.7
Other Durables	2.8*	6.0	2.9*	4.0	46.5*	45.1	7.9	17.0	24.3	16.8	18.1	27.9
Non-Durables												
Food	6.4*	6.3	7.4	3.1	35.1*	45.5	13.0*	20.0	24.4*	23.5	24.2	18.1
Textiles, Printing, Other Non-Durables	8.1*	11.3	2.5*	2.3	49.2	34.8	10.2	6.8	23.5	31.6	17.5*	15.6

(continued)

TABLE 18

Detroit SMSA, 1970
Comparison Between Census 2% and EEOC Industry by Occupation
Percent Female and Percent Negro [1]

	Occupation											
	Prof. Tech. Mgrs.				Sales, Service, Clerk				Craft. Oper. Laborer			
	% Female		% Negro		% Female		% Negro		% Female		% Negro	
Industry	EEOC	Census	EEOC	Census	EEOC	Census	EEOC	Census	EEOC	Census	EEOC	Census
Non Manufacturing												
Construction	1.2	8.1	2.3*	4.3	51.6*	54.8	10.5*	9.0	0.2	1.2	10.9	14.0
Transport. Comm.	10.7*	12.7	12.7	4.2	65.1*	61.8	11.3	16.3	2.1	6.1	7.0	15.5
Wholesale	4.2*	6.5	2.8*	3.9	41.0	34.3	7.7	5.7	6.4*	7.6	23.3	19.1
Retail	25.9*	24.6	4.3*	5.8	67.2*	69.9	12.9*	11.8	18.9	12.7	17.6	10.5
Finance, Ins. Real Estate	16.4	21.7	4.6	7.9	76.5	62.4	16.1	12.2	23.1*	23.4	24.2*	19.1
Service Non Profit	37.6	50.7	12.0*	11.1	65.1	77.0	36.9	25.7	20.5	27.0	47.2	26.8

[1] Excluding agriculture, public administration and private household. Mining not shown due to possible disclosure.

*Estimate from EEOC within two standard errors of Census estimate.

TABLE 19

MINORITY GROUP EMPLOYMENT BY OCCUPATION AND SEX FOR SELECTED INDUSTRIES AND STANDARD METROPOLITAN STATISTICAL AREAS, 1967

U. S. EQUAL EMPLOYMENT OPPORTUNITY COMMISSION

	TOTAL EMPLOY-MENT	WHITE COLLAR OCCUPATIONS						BLUE COLLAR OCCUPATIONS				SERVICE WORKERS
		TOTAL WHITE COLLAR EM-PLOYMENT	OFFICIALS AND MANAGERS	PROFESS-IONALS	TECHNI-CIANS	SALES WORKERS	OFFICE AND CLERICAL	TOTAL BLUE COLLAR EM-PLOYMENT	CRAFTS-MEN	OPER-ATIVES	LABORERS	
DENVER, COLORADO												
GENERAL MERCHANDISE STORES		SIC 53										
SPANISH SURNAMED AMERICAN												
MALE #	190	74	14	0	7	41	12	76	10	49	17	40
OCCUPATIONAL DISTR.	100.0	38.9	7.4	.0	3.7	21.6	6.3	40.0	5.3	25.8	8.9	21.1
PARTICIPATION RATE	4.5	2.7	1.4	.0	7.4	2.8	6.4	7.9	4.1	9.1	9.3	8.5
FEMALE #	421	293	7	1	3	187	95	41	7	23	11	87
OCCUPATIONAL DISTR.	100.0	69.6	1.7	.2	.7	44.4	22.6	9.7	1.7	5.5	2.6	20.7
PARTICIPATION RATE	5.1	4.2	1.6	2.8	7.5	4.3	4.4	8.4	8.0	7.0	14.7	12.3
TOTAL #	611	367	21	1	10	228	107	117	17	72	28	127
OCCUPATIONAL DISTR.	100.0	60.1	3.4	.2	1.6	37.3	17.5	19.1	2.8	11.8	4.6	20.8
PARTICIPATION RATE	4.9	3.7	1.5	1.1	7.5	3.9	4.6	8.0	5.1	8.3	10.9	10.8
NEGRO												
MALE #	211	49	10	0	1	31	7	39	6	24	9	123
OCCUPATIONAL DISTR.	100.0	23.2	4.7	.0	.5	14.7	3.3	18.5	2.8	11.4	4.3	58.3
PARTICIPATION RATE	5.0	1.8	1.0	.0	1.1	2.1	3.7	4.0	2.4	4.5	4.9	20.2
FEMALE #	147	87	2	0	1	66	18	3	1	1	1	57
OCCUPATIONAL DISTR.	100.0	59.2	1.4	.0	.7	44.9	12.2	2.0	.7	.7	.7	38.8
PARTICIPATION RATE	1.8	1.2	.4	.0	2.5	1.5	.8	.6	1.1	.3	1.3	8.0
TOTAL #	358	136	12	0	2	97	25	42	7	25	10	180
OCCUPATIONAL DISTR.	100.0	38.0	3.4	.0	.6	27.1	7.0	11.7	2.0	7.0	2.8	50.3
PARTICIPATION RATE	2.9	1.4	.8	.0	1.5	1.7	1.1	2.9	2.1	2.9	3.9	15.3
ALL EMPLOYEES												
MALE #	4204	2770	980	53	94	1455	188	965	246	537	182	469
OCCUPATIONAL DISTR.	100.0	65.9	23.3	1.3	2.2	34.6	4.5	23.0	5.9	12.8	4.3	11.2
FEMALE #	8253	7055	449	36	40	4388	2142	489	87	327	75	709
OCCUPATIONAL DISTR.	100.0	85.5	5.4	.4	.5	53.2	26.0	5.9	1.1	4.0	.9	8.6
TOTAL (44 ESTAB.)	12457	9825	1429	89	134	5843	2330	1454	333	864	257	1178
OCCUPATIONAL DISTR.	100.0	78.9	11.5	.7	1.1	46.9	18.7	11.7	2.7	6.9	2.1	9.5
FOOD STORES		SIC 54										
SPANISH SURNAMED AMERICAN												
MALE #	296	157	14	5	0	137	1	95	27	56	12	44
OCCUPATIONAL DISTR.	100.0	53.0	4.7	1.7	.0	46.3	.3	32.1	9.1	18.9	4.1	14.9
PARTICIPATION RATE	4.8	4.8	1.7	4.7	.0	6.2	.9	5.4	4.9	5.1	12.2	3.7
FEMALE #	66	35	0	0	0	26	9	30	0	23	7	1
OCCUPATIONAL DISTR.	100.0	53.0	.0	.0	.0	39.4	13.6	45.5	.0	34.8	10.6	1.5
PARTICIPATION RATE	3.5	2.7	.0	.0	.0	3.0	2.2	9.1	.0	10.7	9.6	.4
TOTAL #	362	192	14	5	0	163	10	125	27	79	19	45
OCCUPATIONAL DISTR.	100.0	53.0	3.9	1.4	.0	45.0	2.8	34.5	7.5	21.8	5.2	12.4
PARTICIPATION RATE	4.5	4.2	1.6	4.5	.0	5.3	1.9	6.0	4.6	6.0	11.1	3.1
NEGRO												
MALE #	193	50	1	1	0	47	1	50	4	33	13	93
OCCUPATIONAL DISTR.	100.0	25.9	.5	.5	.0	24.4	.5	25.9	2.1	17.1	6.7	48.2
PARTICIPATION RATE	3.1	1.5	.1	.9	.0	2.1	.9	2.8	.7	3.0	13.3	7.9
FEMALE #	41	29	0	0	0	23	6	3	2	1	0	9
OCCUPATIONAL DISTR.	100.0	70.7	.0	.0	.0	56.1	14.6	7.3	4.9	2.4	.0	22.0
PARTICIPATION RATE	2.2	2.2	.0	.0	.0	2.6	1.5	.9	4.8	.5	.0	3.2
TOTAL #	234	79	1	1	0	70	7	53	6	34	13	102
OCCUPATIONAL DISTR.	100.0	33.8	.4	.4	.0	29.9	3.0	22.6	2.6	14.5	5.6	43.6
PARTICIPATION RATE	2.9	1.7	.1	.9	.0	2.3	1.4	2.5	1.0	2.6	7.6	7.0
ALL EMPLOYEES												
MALE #	6200	3270	836	106	4	2213	111	1755	549	1108	98	1175
OCCUPATIONAL DISTR.	100.0	52.7	13.5	1.7	.1	35.7	1.8	28.3	8.9	17.9	1.6	19.0
FEMALE #	1904	1292	13	6	0	868	405	329	42	214	73	283
OCCUPATIONAL DISTR.	100.0	67.9	.7	.3	.0	45.6	21.3	17.3	2.2	11.2	3.8	14.9
TOTAL (14 ESTAB.)	8104	4562	849	112	4	3081	516	2084	591	1322	171	1458
OCCUPATIONAL DISTR.	100.0	56.3	10.5	1.4	.0	38.0	6.4	25.7	7.3	16.3	2.1	18.0
AUTOMOTIVE DEALERS		SIC 55										
SPANISH SURNAMED AMERICAN												
MALE #	24	1	0	0	0	1	0	23	6	5	12	0
OCCUPATIONAL DISTR.	100.0	4.2	.0	.0	.0	4.2	.0	95.8	25.0	20.8	50.0	.0
PARTICIPATION RATE	3.9	.3	.0	.0	.0	.5	.0	7.3	5.9	3.1	22.2	.0
FEMALE #	1	1	0	0	0	1	0	0	0	0	0	0
OCCUPATIONAL DISTR.	100.0	100.0	.0	.0	.0	100.0	.0	.0	.0	.0	.0	.0
PARTICIPATION RATE	1.7	1.7	.0	.0	.0	33.3	.0	.0	.0	.0	.0	.0
TOTAL #	25	2	0	0	0	2	0	23	6	5	12	0
OCCUPATIONAL DISTR.	100.0	8.0	.0	.0	.0	8.0	.0	92.0	24.0	20.0	48.0	.0
PARTICIPATION RATE	3.7	.6	.0	.0	.0	1.0	.0	7.3	5.9	3.1	21.8	.0

MINORITY GROUP EMPLOYMENT BY OCCUPATION AND SEX FOR SELECTED INDUSTRIES AND STANDARD METROPOLITAN STATISTICAL AREAS, 1967
U. S. EQUAL EMPLOYMENT OPPORTUNITY COMMISSION

	TOTAL EMPLOYMENT	WHITE COLLAR OCCUPATIONS						BLUE COLLAR OCCUPATIONS				SERVICE WORKERS
		TOTAL WHITE COLLAR EMPLOYMENT	OFFICIALS AND MANAGERS	PROFESSIONALS	TECHNICIANS	SALES WORKERS	OFFICE AND CLERICAL	TOTAL BLUE COLLAR EMPLOYMENT	CRAFTSMEN	OPERATIVES	LABORERS	
DENVER, COLORADO												
AUTOMOTIVE DEALERS		SIC 55			(CONT)							
NEGRO												
MALE #	28	2	1	0	0	1	0	26	3	8	15	0
OCCUPATIONAL DISTR.	100.0	7.1	3.6	.0	.0	3.6	.0	92.9	10.7	28.6	53.6	.0
PARTICIPATION RATE	4.5	.7	1.2	.0	.0	.5	.0	8.2	3.0	5.0	27.8	.0
FEMALE #	0	0	0	0	0	0	0	0	0	0	0	0
OCCUPATIONAL DISTR.	100.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
PARTICIPATION RATE	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
TOTAL #	28	2	1	0	0	1	0	26	3	8	15	0
OCCUPATIONAL DISTR.	100.0	7.1	3.6	.0	.0	3.6	.0	92.9	10.7	28.6	53.6	.0
PARTICIPATION RATE	4.1	.6	1.1	.0	.0	.5	.0	8.3	3.0	5.0	27.3	.0
ALL EMPLOYEES												
MALE #	618	300	86	1	0	205	8	316	101	161	54	2
OCCUPATIONAL DISTR.	100.0	48.5	13.9	.2	.0	33.2	1.3	51.1	16.3	26.1	8.7	.3
FEMALE #	59	58	2	0	0	3	53	1	0	0	1	0
OCCUPATIONAL DISTR.	100.0	98.3	3.4	.0	.0	5.1	89.8	1.7	.0	.0	1.7	.0
TOTAL (14 ESTAB.)	677	358	88	1	0	208	61	317	101	161	55	2
OCCUPATIONAL DISTR.	100.0	52.9	13.0	.1	.0	30.7	9.0	46.8	14.9	23.8	8.1	.3
APPAREL & ACCESSORY STORES		SIC 56										
SPANISH SURNAMED AMERICAN												
MALE #	25	23	3	0	0	20	0	0	0	0	0	2
OCCUPATIONAL DISTR.	100.0	92.0	12.0	.0	.0	80.0	.0	.0	.0	.0	.0	8.0
PARTICIPATION RATE	6.6	7.3	3.0	.0	.0	10.5	.0	.0	.0	.0	.0	3.6
FEMALE #	13	13	0	0	0	7	6	0	0	0	0	0
OCCUPATIONAL DISTR.	100.0	100.0	.0	.0	.0	53.8	46.2	.0	.0	.0	.0	.0
PARTICIPATION RATE	2.9	3.5	.0	.0	.0	5.0	3.0	.0	.0	.0	.0	.0
TOTAL #	38	36	3	0	0	27	6	0	0	0	0	2
OCCUPATIONAL DISTR.	100.0	94.7	7.9	.0	.0	71.1	15.8	.0	.0	.0	.0	5.3
PARTICIPATION RATE	4.6	5.2	2.3	.0	.0	8.2	2.8	.0	.0	.0	.0	2.9
NEGRO												
MALE #	20	6	0	2	0	4	0	0	0	0	0	14
OCCUPATIONAL DISTR.	100.0	30.0	.0	10.0	.0	20.0	.0	.0	.0	.0	.0	70.0
PARTICIPATION RATE	5.3	1.9	.0	25.0	.0	2.1	.0	.0	.0	.0	.0	25.0
FEMALE #	20	10	0	0	0	3	7	2	0	2	0	8
OCCUPATIONAL DISTR.	100.0	50.0	.0	.0	.0	15.0	35.0	10.0	.0	10.0	.0	40.0
PARTICIPATION RATE	4.5	2.7	.0	.0	.0	2.1	3.4	3.7	.0	12.5	.0	66.7
TOTAL #	40	16	0	2	0	7	7	2	0	2	0	22
OCCUPATIONAL DISTR.	100.0	40.0	.0	5.0	.0	17.5	17.5	5.0	.0	5.0	.0	55.0
PARTICIPATION RATE	4.9	2.3	.0	25.0	.0	2.1	3.3	3.4	.0	12.5	.0	32.4
ALL EMPLOYEES												
MALE #	376	316	100	8	5	191	12	4	4	0	0	56
OCCUPATIONAL DISTR.	100.0	84.0	26.6	2.1	1.3	50.8	3.2	1.1	1.1	.0	.0	14.9
FEMALE #	442	376	33	0	0	140	203	54	6	16	32	12
OCCUPATIONAL DISTR.	100.0	85.1	7.5	.0	.0	31.7	45.9	12.2	1.4	3.6	7.2	2.7
TOTAL (21 ESTAB.)	818	692	133	8	5	331	215	58	10	16	32	68
OCCUPATIONAL DISTR.	100.0	84.6	16.3	1.0	.6	40.5	26.3	7.1	1.2	2.0	3.9	8.3
EATING & DRINKING PLACES		SIC 58										
SPANISH SURNAMED AMERICAN												
MALE #	84	0	0	0	0	0	0	5	0	5	0	79
OCCUPATIONAL DISTR.	100.0	.0	.0	.0	.0	.0	.0	6.0	.0	6.0	.0	94.0
PARTICIPATION RATE	16.4	.0	.0	.0	.0	.0	.0	21.7	.0	55.6	.0	18.3
FEMALE #	44	0	0	0	0	0	0	0	0	0	0	44
OCCUPATIONAL DISTR.	100.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	100.0
PARTICIPATION RATE	6.5	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	7.0
TOTAL #	128	0	0	0	0	0	0	5	0	5	0	123
OCCUPATIONAL DISTR.	100.0	.0	.0	.0	.0	.0	.0	3.9	.0	3.9	.0	96.1
PARTICIPATION RATE	10.8	.0	.0	.0	.0	.0	.0	21.7	.0	55.6	.0	11.6
NEGRO												
MALE #	115	0	0	0	0	0	0	6	1	4	1	109
OCCUPATIONAL DISTR.	100.0	.0	.0	.0	.0	.0	.0	5.2	.9	3.5	.9	94.8
PARTICIPATION RATE	22.5	.0	.0	.0	.0	.0	.0	26.1	10.0	44.4	25.0	25.3
FEMALE #	65	1	0	0	0	0	1	0	0	0	0	64
OCCUPATIONAL DISTR.	100.0	1.5	.0	.0	.0	.0	1.5	.0	.0	.0	.0	98.5
PARTICIPATION RATE	9.7	[illegible]	.0	.0	.0	.0	3.3	.0	.0	.0	.0	10.2
TOTAL #	180	1	0	0	0	0	1	6	1	4	1	173
OCCUPATIONAL DISTR.	100.0	.6	.0	.0	.0	.0	.6	3.3	.6	2.2	.6	96.1
PARTICIPATION RATE	15.2	1.0	.0	.0	.0	.0	3.1	26.1	10.0	44.4	25.0	16.4

BIBLIOGRAPHY

1) Arets, Paul, "Summary of the Attitudes and Needs of Employers and their Organizations in Regard to Public Employment Services," in *The Public Employment Services and Management*, International Management Services, Organiations for Economic Co-operation and Development, 1966.

2) Chavrid, Vladimir D., "Review of Emerging Needs and Expected Changes in the Labor Market Information Program," remarks to the first National Conference of State Employment Security Chiefs of Research and Analysis, Denver, Colorado, October 15, 1970.

3) Childs, D. L., STIS Technical Report 1 - Development of a Set Theoretic Data Structure, Set-Theoretic Information Systems Corp., Ann Arbor, Michigan, p.3.

4) Childs, D.L., "Feasibility of a Set-Theoretic Data Structure - A General Structure Based on a Reconstituted Definition of Relation," IFIP Congress, Proc. 1968, pp. 162-172.

5) Childs, D.L., "Description of a Set-Theoretic Data Structure," ACM Proc. 1968 FJCO, Vol. 33 Part 1, pp. 557-564.

6) Cooperative Area Manpower Planning Committee, *State of Michigan Cooperative Manpower Plan for Fiscal 1970*, 1970, pp. 127-299.

7) Goldstein, Harold, "Data Sources for State and Local Manpower Planning," remarks to the National Conference on State and Local Manpower Policy Planning, Salt Lake City, Utah, April 25, 1971.

8) Hershey, William R., "A User's Guide to the MICRO Computer Language," Institute of Labor and Industrial Relations, University of Michigan.

9) Holt, Charles C., "How Can the Phillips' Curve Be Moved to Reduce Both Inflation and Unemployment?" in Edmund S. Phelps, *et. al.*, *Microeconomic Foundations of Employment and Inflation Theory*, New York: W. W. Norton & Co., 1968, pp. 224-256.

10) Jobin, P. B., and Stern, J. R., "The Shift from Blue-Collar to White-Collar Jobs," *Monthly Labor Review*, October, 1969, pp. 7-13.

11) Kalacheck, Edward, "The Youth Labor Market," *Policy Papers in Industrial Relations #12*, The Institue of Labor and Industrial Relations, University of Michigan and Wayne State University, 1969.

12) Lawson, Rodger S., "Perspectives on the Development of a Comprehensive Labor Market Information System," *Methods for Manpower Analysis*, W. E. Upjohn Institute, April, 1973.

13) Levine, Louis, and Norton, J.H., Manpower Information for Urban Poverty Areas, The Pennsylvania State University Institute for Research on Human Resources, University Park, Penn., 1969, pp. 1-15, 48-66, 72-95, and abstract.

14) STIS Corp., STDS-1 Operations Manual, 117 North First Street, Ann Arbor, Michigan, December, 1971.

15) Thal-Larsen, Margaret T., "Requirements and Design of a Labor Market Information for a Large Metropolitan Area," Berkely, 1973.

16) U.S. Bureau of Labor Statistics, Employment and Earnings, Vol. 17, No. 11, May, 1971.

17) U.S. President, Manpower Report of the President, 1971, p. 68.

18) Wayne State University, Social Reporting in Michigan: Problems and Issues, released by the Office of Planning Co-ordination, Bureau of Policies and Programs, Technical Report A-37, written by the Center for Urban Studies, 1970, pp. 1-72, 95-129, 160-198, 263-281.

19) Wilcock, R. C., and Franke, W. H., Unwanted Workers, Free Press of Glencoe, 1963, pp. 114-129.

20) Yavitz, Boris and Morse, Dean, The Labor Market: An Information System, New York: Praeger, 1973.